Articles

Slouching Towards Equality:
Gender Discrimination, Market Efficiency,
and Social Change

Edward J. McCaffery[†]

CONTENTS

I. INTRODUCTION 596

II. THE BASIC FACTS 600

III. THE BASIC MODEL 602

 A. Preliminary Assumptions 603

 B. Phase I: Rational Discrimination 603

[†] Visiting Associate Professor of Law, Yale Law School (1993-94); Associate Professor of Law, University of Southern California Law Center. I owe thanks to Bruce Ackerman, Scott Altman, Alex Capron, Howard Chang, Dick Craswell, Ron Garet, Tom Griffith, Henry Hansmann, Deborah Hensler, Joni Hersch, Greg Keating, Jeff Lehman, Nancy Marder, Uriel Procaccia, Judith Resnik, Roberta Romano, Carol Rose, Elyn Saks, Stewart Schwab, Michael Shapiro, Larry Simon, Larry Solum, Matt Spitzer, Nomi Stolzenberg, Chris Stone, and Jeff Strnad for their helpful comments, and to all of the participants at Faculty Workshops at the University of Southern California Law Center and Yale Law School, where earlier drafts of the Article were presented. Susan Rindner provided valuable research assistance, and the library staff at the U.S.C. Law Center was, as always, outstanding. I would also like to thank Susette Passaro, Kirk Stark, and all of the other students on the *Yale Law Journal* who worked on the Article. After a draft of this Article was substantially completed, I received a very gracious and detailed letter from Richard Epstein, listing numerous criticisms. I have tried to respond in this final version to many of Professor Epstein's comments, although substantial disagreements remain on many points.

- 1. *Imperfect Contracts* 606
- 2. *Incomplete and Asymmetric Information* 608
- C. *Phase II: Self-Fulfilling Prophecy* 615
 - 1. *Taxes* 617
 - 2. *Incomplete Markets* 619
 - 3. *Human Capital Decisions* 622
- D. *Phase III: Slouching Towards Equality* 624
 - 1. *Separating Equilibria* 626
 - 2. *Signaling Behavior* 630
- E. *Summary of Model* 633
- IV. THE TROUBLE WITH DOING NOTHING 634
 - A. *What Is Efficiency?* 636
 - B. *Why Do We Care About Efficiency?* 639
 - 1. *Efficiency and Utilitarianism* 640
 - 2. *Efficiency and Liberalism* 643
 - 3. *Efficiency and Libertarianism* 646
 - C. *The Case Against Doing Nothing* 648
 - D. *Why Redistribution Is Not Enough* 651
 - E. *Equality of What?* 652
- V. GROPING FOR SOLUTIONS 654
 - A. *Optimal Tax Theory* 657
 - B. *Returning to the Model* 664
 - C. *The Normative Propriety of the Proposed Solution* 666
 - D. *Extensions* 672
- VI. CONCLUSION 673

I. INTRODUCTION

There are two overlapping debates regarding gender discrimination in the workplace. The first of these—what I shall call the descriptive debate—is an argument over the direction and magnitude of actual change. The pessimists, gaining prominence in the popular literature, maintain that there has been a turn in the march toward greater equality—a backlash against the feminist movement.¹ Opponents of this camp brandish a variety of statistics, mostly

1. See SUSAN FALUDI, BACKLASH: THE UNDECLARED WAR AGAINST AMERICAN WOMEN (1991); ARLIE HOCHSCHILD, THE SECOND SHIFT: WORKING PARENTS AND THE REVOLUTION AT HOME (1989); Madeline E. Heilman et al., *Has Anything Changed? Current Characterizations of Men, Women, and Managers*, 74 J. APPLIED PSYCHOL. 935 (1989); Joan Z. Spade & Carole A. Reese, *We've Come a Long Way, Maybe: College Students' Plans for Work and Family*, 24 SEX ROLES 309 (1991).

related to the gender wage gap, to counter that the long-term trend is indeed toward equality. Even if the movement has slowed down recently, the optimists believe that we are heading, more or less inexorably, toward a better day.²

The second debate—what I shall call the interventionist debate—is over the propriety and efficacy of legal remedies to reduce discrimination. The anti-interventionists, exemplified by Richard Posner and Richard Epstein, argue that legal interventions are virtually always inefficient, and that the market will naturally lead to whatever is the efficient level of discrimination, if any.³ In contrast, the interventionists maintain either that there is insufficient reason to believe in the efficiency of the market, or that efficiency is an inadequate goal for addressing all of the concerns over gender discrimination.⁴

Advocates on both sides of these debates make predictable arguments. Anti-interventionists tend to stress the optimistic strand in the descriptive debate, arguing that legal remedies have, more often than not, impeded the progress toward the promised land. Interventionists, on the other hand, are more likely to stress the pessimistic strand in the descriptive debate, maintaining that whatever positive trend exists has been due to, or at least aided by, prior interventions.

I want to change the terms of these debates. There is some truth in all of the above positions, and recognizing this will reveal new perspectives on

2. See, e.g., FRANCINE D. BLAU & MARIANNE A. FERBER, *THE ECONOMICS OF WOMEN, MEN, AND WORK* 132-38 (2d ed. 1992); CLAUDIA GOLDIN, *UNDERSTANDING THE GENDER GAP* 215-17 (1990); FABIAN LINDEN, *MULTI-PAYCHECK FAMILIES* 3-5, 12 (1993); ELAINE SORENSSEN, *EXPLORING THE REASONS BEHIND THE NARROWING GENDER GAP* (1991); Judith Fields & Edward N. Wolff, *The Decline of Sex Segregation and the Wage Gap, 1970-80*, 26 J. HUM. RESOURCES 608 (1991); Diane Crispell, *Women's Earnings Gap Is Closing—Slowly*, AM. DEMOGRAPHICS, Feb. 1991, at 14; Michael W. Horrigan & James P. Markey, *Recent Gains in Women's Earnings: Better Pay or Longer Hours?*, MONTHLY LAB. REV., July 1990, at 11; Sylvia Nasar, *Women's Progress Stalled? Just Not So*, N.Y. TIMES, Oct. 18, 1992, at C1.

3. See RICHARD A. EPSTEIN, *FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS* (1992); Richard A. Posner, *An Economic Analysis of Sex Discrimination Laws*, 56 U. CHI. L. REV. 1311 (1989) [hereinafter, Posner, *Economic Analysis*]; Richard A. Posner, *The Efficiency and Efficacy of Title VII*, 136 U. PA. L. REV. 513 (1987) [hereinafter Posner, *Efficiency and Efficacy*]; Daniel R. Fischel & Edward P. Lazear, *Comparable Worth and Discrimination in Labor Markets*, 53 U. CHI. L. REV. 891 (1986). Professor Epstein points out that he is the only author to call explicitly for repeal of the antidiscrimination laws, although I believe that we both agree that this position is at least implicit in Posner's work. Letter from Richard A. Epstein, Professor of Law, University of Chicago Law School, to Edward J. McCaffery, Visiting Associate Professor of Law, Yale Law School (June 23, 1993) (on file with author).

4. For interventionist arguments in the case of gender discrimination, see John J. Donohue III, *Prohibiting Sex Discrimination in the Workplace: An Economic Perspective*, 56 U. CHI. L. REV. 1337 (1989) [hereinafter Donohue, *Economic Perspective*]; John J. Donohue III, *Is Title VII Efficient?*, 134 U. PA. L. REV. 1411 (1986) [hereinafter Donohue, *Title VII*]; Mary E. Becker, *Barriers Facing Women in the Wage-Labor Market and the Need for Additional Remedies: A Reply to Fischel and Lazear*, 53 U. CHI. L. REV. 934 (1986); James D. Holzhauser, *The Economic Possibilities of Comparable Worth*, 53 U. CHI. L. REV. 919 (1986). For more direct assaults on the role of efficiency or other market-based analysis, see Marjorie E. Kornhauser, *The Rhetoric of the Anti-Progressive Income Tax Movement: A Typical Male Reaction*, 86 MICH. L. REV. 465 (1987); Frances E. Olsen, *The Family and the Market: A Study of Ideology and Legal Reform*, 96 HARV. L. REV. 1497 (1983); Margaret J. Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849 (1987).

gender discrimination. As to the descriptive debate, I argue that it is not just whether and when the gender gap⁵ narrows, by approaching pay equity for women, but also *how* it narrows.⁶ Much of the literature on and popular awareness of gender discrimination has focused on the gender wage gap and other objective indicia such as relative participation rates in different occupations. But there are compelling reasons to believe that these are dangerously wrong measures of what truly matters. As I hope to demonstrate, today's labor markets are a product of economic distortions created in an era of male domination of labor. Measuring equality according to terms appropriated from that era may be misleading or worse. This central point helps to explain why the statistics that suggest improvement are out of synch with the perception of retrenchment. We are slouching towards a predefined notion of equality.⁷

5. I use "gender gap" in the same sense that GOLDIN, *supra* note 2, and SORESENSEN, *supra* note 2, and others do, namely that of a gap between the salaries of men and women. For other discussions of the gender gap, see Francine D. Blau & Lawrence M. Kahn, *The Gender Earnings Gap: Learning From International Comparisons*, 82 AM. ECON. REV. (PAPERS & PROC.) 533 (1992); Pamela J. Loprest, *Gender Differences in Wage Growth and Job Mobility*, 82 AM. ECON. REV. (PAPERS & PROC.) 526 (1992). Generally, this earnings gap is aggregated across different professions, so that it captures some of the effect of women's concentration in lower paying fields; at other times, it is industry-specific.

6. It is not just the quantitative measure of equality which is relevant, but also its quality. As Amartya Sen has argued, all persuasive political theories are egalitarian in some sense; the crucial question is the choice of variable(s) against which to measure equality. The story of this Article becomes, in part, the story of the consequences of focusing on the wrong type of equality. For a discussion of these issues, see AMARTYA K. SEN, *INEQUALITY REEXAMINED* (1992).

7. The theme that the work force is defined in largely male terms is well developed in the feminist literature. See, e.g., BARBARA R. BERGMANN, *THE ECONOMIC EMERGENCE OF WOMEN* 4-5 (1986); FELICE N. SCHWARTZ, *BREAKING WITH TRADITION: WOMEN AND WORK, THE NEW FACTS OF LIFE* (1992); Kathryn Abrams, *Gender Discrimination and the Transformation of Workplace Norms*, 42 VAND. L. REV. 1183 (1989); Karen Czapsanskiy, *Volunteers and Draftees: The Struggle for Parental Equality*, 38 UCLA L. REV. 1415 (1991); Nancy E. Dowd, *Work and Family: Restructuring the Workplace*, 32 ARIZ. L. REV. 431 (1990). The broader theme that equality talk often obscures other important concerns is also well developed. See, e.g., *FEMINIST LEGAL THEORY*, parts 2 and 3 (D. Kelly Weisberg ed., 1993); Lucinda M. Finley, *Transcending Equality Theory: A Way Out of the Maternity and the Workplace Debate*, 86 COLUM. L. REV. 1118 (1986); Nadine Taub & Wendy W. Williams, *Will Equality Require More than Assimilation, Accommodation or Separation from the Existing Social Structure?*, 37 RUTGERS L. REV. 825 (1985). One of my goals in this Article is to provide an economic explanation for these viewpoints. In doing so, I do not intend to take on directly the "sameness versus difference" debate about whether or not women are the "same" as men. For a fuller discussion of those issues, see Joan C. Williams, *Deconstructing Gender*, 87 MICH. L. REV. 797 (1989). My hope is to create the conditions under which autonomous preferences can flourish, and not to make any a priori characterization along essentialist lines. It may seem that I am taking a different tack, with a focus on the fact that women have been differently affected by labor force dynamics than men, and my belief—a large part of the inspiration for the project—that the work force *ought* to be changing more than it is to accommodate an influx of married women. But this is not necessarily so. As to the first point, I shall argue that women can be differentially affected even if they have exactly the same preferences regarding work and family choices as men. As to the second point, I believe that, in a reformulated workplace, men as well as women will take on new roles. In general, I have some affinity with Catharine MacKinnon, who has argued that power is the important variable, not sameness or difference. See, e.g., CATHARINE A. MACKINNON, *FEMINISM UNMODIFIED: DISCOURSES ON LIFE AND LAW* 51 (1987) ("Gender here is a matter of dominance, not difference. . . . Explaining the subordination of women to men, a political condition, has nothing to do with difference in any fundamental sense."). Current biases take power away from women.

As to the interventionist debate, I concede that efficiency is an important goal, even a necessary condition for any attractive social change, but it is crucial to ground efficiency analysis in an appealing normative framework. When we consider a broad range of market failures and their effects over time against a normatively appealing version of efficiency, the case for non-intervention weakens, perhaps fatally. In particular, informational problems and the present tax system interact to create significant barriers to women's participation in the work force. As a result, women find that they face a stark choice: prove that they can act like men have always acted or stay home. We are left with only two basic models of the dual-spouse family: either there are two full-time workers or the wife stays home. This normative framework allows us to conclude that, even if the optimists are right that the gender gap is narrowing and the anti-interventionists are right that efficiency is an important norm to pursue and that non-intervention will allow the market to reach an efficient equilibrium, we still ought to consider quite carefully the case for various governmental actions on efficiency grounds. As this Article demonstrates, this analysis leads to recommended actions that differ from those proposed in the general run of the antidiscrimination literature. In particular, I propose, at least for illustrative and discussion purposes, a reform relying less on regulatory interventions such as Title VII and more on fundamental change in the tax laws. This reform accepts part but by no means all of the arguments against regulation, and suggests alternative strong social action to address gender discrimination.

This description of my project indicates its primary goals. I mean to show that arguments from economic efficiency, properly understood, do not lead to the do-nothing conclusions of Epstein, Posner, and others. I hope to show, particularly to those who are skeptical of all law and economics claims, the possible appeal of efficiency analysis. And I hope to underscore the importance of focusing on a rich and attractive sense of "equality," rather than on static, objective measures such as the gender wage gap. In the end, many will still reject all forms of economic thought or even liberalism as being insufficient to meet the task of eradicating sexism (and other social ills), but there is value in stating the strongest case for such a philosophy before rejecting it out of hand.

The Article is divided into the following Parts. Part II sets out some basic facts for the rest of the analysis. Part III presents a model of gender discrimination in the work force and extends it through three dynamic phases. Here I employ simple mathematical, graphic, and illustrative techniques, relegating technical discussions to the notes. Part IV explores the meaning and propriety of efficiency and then uses the model to criticize the laissez-faire approach of Posner, Epstein, and others. Part V outlines an illustrative reform proposal based on a fundamental change in the tax laws and a lessening of the

present regulatory intervention. Part VI wraps up with a summary and some conclusions.

II. THE BASIC FACTS

Before progressing to the model of gender discrimination in the work force, I want to set forth the facts that animate my analysis. I have deliberately attempted to rely on the least controversial empirical data relating to gender discrimination. I do not intend to convince the reader of my particular view of the relevant statistics, but rather to rely only on widely accepted data. The success of the model as an explanatory device depends on its mapping with these basic facts.

One fact that seems undeniable is that the gender gap has been narrowing, moving towards elimination over the past century.⁸ Progress has not been steady; periods of significant improvement have been followed by periods of inertia or even retrenchment.⁹ Although the clear trend is toward closing the gap, debate remains as to when, if at all, it will close completely. Almost all of the existing wage gap is between married men and ever-married women; remove them from the analysis, and women and men receive virtually equal pay.¹⁰ The gap is, and always has been, largely explainable in terms of job market segregation.¹¹ Because women have been disproportionately represented in lower-paying sectors of the economy, female earnings have always been lower than male wages.¹² The important role of job market segregation, however, should not overshadow the impact of wage discrimination within particular job categories in perpetuating the gender gap.¹³ In particular, women have failed to reach the upper echelons of hierarchies in a wide range of fields.¹⁴

8. See sources cited *supra* note 2.

9. See GOLDIN, *supra* note 2, at 62; SORESENSEN, *supra* note 2, at 15, 17.

10. See, e.g., Yinon Cohen & Yitchak Haberfeld, *Why Do Married Men Earn More than Unmarried Men?*, 20 SOC. SCI. RES. 29 (1991); Fischel & Lazear, *supra* note 3, at 899 ("Women who have never been married appear to have complete wage parity with men who have never been married."). But cf. Martha S. Hill, *The Wage Effects of Marital Status and Children*, 14 J. HUM. RESOURCES 579, 592 (1979) (identifying marriage's large positive effect on men's wages, but no noticeable detrimental effect on women's wages). The case seems to have changed since 1979, however, so that both effects are now pronounced. Throughout the Article, "ever-married" refers to women who have once been married, whether they are now still married, divorced, separated, or widowed.

11. See BLAU & FERBER, *supra* note 2; GOLDIN, *supra* note 2, at 75 ("Occupational segregation by sex has been and still is extensive. Although the data remain fragmented, it is likely that occupational segregation by sex was more extensive around the turn of the century than it is today and that it declined at various points, particularly in recent decades."); SORESENSEN, *supra* note 2; Carol Rose, *Women and Property: Gaining and Losing Ground*, 78 VA. L. REV. 421, 436 (1992) (citing, *inter alia*, JERRY A. JACOBS, *REVOLVING DOORS: SEX SEGREGATION AND WOMEN'S CAREERS* (1989)).

12. For a vivid example of the early role of job market segregation, see GOLDIN, *supra* note 2, at 78 (Table 3.4) (listing "Male-Intensive" and "Female-Intensive" industries for 1890).

13. See generally *id.* at 83-118 (discussing emergence of "wage discrimination").

14. See BLAU & FERBER, *supra* note 2, at 122-26; GOLDIN, *supra* note 2, at 4-5; Ellen Auster & Robert Drazin, *SEX INEQUALITY AT HIGHER LEVELS IN THE HIERARCHY* 216; Thomas A. DiPrete &

Other important trends have accompanied the narrowing of the gender gap. A massive number of women, especially married women, have entered the paid labor force.¹⁵ As the gap has been closing, women have been educating themselves more, both in absolute terms and as compared to men.¹⁶ The turnover rate among women—representing the percentage of women who leave the work force in any given year—has shrunk dramatically over the past fifty years, going from nearly five times that of men to just over twice the amount.¹⁷ The current work-age profile for women—representing the participation rate across the life cycle—has almost exactly the same shape as for men. Like men, women enter the work force in their early twenties and stay there until their sixties.¹⁸ Fertility and marriage rates have dropped along with increases in education and wages.¹⁹

While these changes have been taking place, the structure of work and typical work-family dynamics have not changed very much at all.²⁰ Full-time work remains the norm. Most of the increase in part-time work over the last twenty years has been involuntary, and women of childbearing age are actually *less* likely to be working part-time today than they were twenty or thirty years ago.²¹ Women are still overwhelmingly likely to be the secondary, or lesser, earner in two-earner families, although the incidence of female-headed, single-earner households has risen dramatically.²² While married women have been

Whitman T. Soule, *Gender and Promotion in Segmented Job Ladder Systems*, 53 AM. SOC. REV. 26 (1988); Edward Lazear & Sherwin Rosen, *Male-Female Wage Differentials in Job Ladders*, 8 J. LAB. ECON. 106 (1990).

15. See U.S. DEP'T OF COMMERCE, BUREAU OF THE CENSUS, STATISTICAL ABSTRACT OF THE UNITED STATES 387-88 (Table Nos. 618, 620) (112th ed. 1992) [hereinafter STATISTICAL ABSTRACT]; Morley Gunderson, *Male-Female Wage Differentials and Policy Responses*, 27 J. ECON. LITERATURE 46 (1989), and sources cited therein.

16. See BLAU & FERBER, *supra* note 2, at 142-46; GOLDIN, *supra* note 2, at 215-17; SORENSEN, *supra* note 2, at 132-33.

17. BLAU & FERBER, *supra* note 2, at 79 (noting that in 1957, female turnover rate was 32.1, compared to 7.8 for males; by 1989, the numbers had become 9.9 and 4.7, respectively); Claudia Goldin, *Monitoring Costs and Occupational Segregation by Sex: A Historical Analysis*, 4 J. LAB. ECON. 1, 24-25 (1986) (discussing, inter alia, dramatic change in female labor market participation after 1950); Audrey Light & Manuelita Ureta, *Panel Estimates of Male and Female Job Turnover Behavior: Can Female Nonquitters Be Identified?*, 10 J. LAB. ECON. 156 (1992); David J. Maume, Jr., *Child-Care Expenditures and Women's Employment Turnover*, 70 SOC. FORCES 495 (1991).

18. BLAU & FERBER, *supra* note 2, at 77-78; STATISTICAL ABSTRACT, *supra* note 15, at 387 (Table No. 618).

19. BLAU & FERBER, *supra* note 2, at 259-64; LINDEN, *supra* note 2, at 4.

20. See BLAU & FERBER, *supra* note 2, at 32; GOLDIN, *supra* note 2, at 211-12; HOCHSCHILD, *supra* note 1, at xi, 12 ("[M]ost workplaces have remained inflexible in the face of family demands of their workers and, at home, most men have yet to really adapt to the changes in the lives of women."); Czapanskiy, *supra* note 7, at 1415; Nancy E. Dowd, *Work and Family: The Gender Paradox and the Limitations of Discrimination Analysis in Restructuring the Workplace*, 24 HARV. C.R.-C.L. L. REV. 79, 84-85 (1989).

21. CHRIS TILLY, *SHORT HOURS, SHORT SHRIFT: CAUSES AND CONSEQUENCES OF PART-TIME WORK* 1 (1990); Martha Chamallas, *Women and Part-Time Work: The Case for Pay Equity and Equal Access*, 64 N.C. L. REV. 709, 711 (1986).

22. GOLDIN, *supra* note 2, at 212; STATISTICAL ABSTRACT, *supra* note 15, at 388 (Table No. 620); Dowd, *supra* note 7, at 97; Howard V. Hayghe, *Family Members in the Work Force*, MONTHLY LAB. REV. March 1990, at 14.

increasingly active in the work force, forty percent of married women with children under six years of age do not work.²³ In contrast, virtually all fathers of young children work.²⁴ Married women, even when they do work full-time jobs in the paid work force, continue to do a vast majority of the work around the home.²⁵

I intend to explain these facts in the model that follows. It is, in sum, a story of growing, quantitative equality, as measured by participation rates and the narrowing of the gender wage gap. As the data demonstrate, however, this equality has been achieved on qualitative terms set in a prior era of male domination of the labor force. The continuing gender division of labor and the absence of more diverse, flexible work arrangements are evidence of the underdeveloped nature of today's labor markets. The question becomes: How did we arrive at such a state?

III. THE BASIC MODEL

This Part sets out a model to explain the development of gender discrimination in U.S. labor markets. I first assume that all discrimination against women in the labor force is rational and then seek to trace the model through several dynamic phases. Section A describes the underlying assumptions of the model, and Sections B through D trace it through three dynamic phases, or temporal periods. Along the way, I stress the relevance of four factors that make existing market structures flawed: (i) transaction costs, (ii) imperfect and asymmetric information, (iii) incomplete markets, and (iv) taxation. Like all models, this one is general and, in important regards, simplistic. I do not intend to give a complete, thick description of gender discrimination. Indeed, one of my central points—borne out by much of the empirical literature—is that such a description may be impossible, due to the numerous possible causes and their complex interactions.²⁶ Instead, using a simple model that broadly fits with existing data, I hope to demonstrate two

23. STATISTICAL ABSTRACT, *supra* note 15, at 388 (Table No. 620).

24. See Dowd, *supra* note 7, at 443, n.78 ("For employed fathers with children under . . . six years of age, their labor participation rate [is] roughly ninety-five percent.").

25. See BLAU & FERBER, *supra* note 2, at 50-56; GOLDIN, *supra* note 2, at 211-12; HOCHSCHILD, *supra* note 1, at 6-10; Rose, *supra* note 11, at 431 (citing, *inter alia*, PAULA ENGLAND & GEORGE FARKAS, HOUSEHOLDS, EMPLOYMENT AND GENDER 94-99 (1986)); see also Joni Hersch, *The Impact of Nonmarket Work on Market Wages*, 81 AM. ECON. REV. (PAPERS & PROC.) 157 (1991); Joni Hersch & Leslie Stratton, *Housework, Effort, and Wages of Married Workers* (1993) (unpublished manuscript, draft on file with author) (noting that greater housework performed by women appears to have a dramatic, negative effect on their market wages).

26. See Glen G. Cain, *The Uses and Limits of Statistical Analysis in Measuring Economic Discrimination*, in ESSAYS ON THE ECONOMICS OF DISCRIMINATION 115 (Emily Hoffman ed., 1991). Almost all discussions of gender discrimination leave open a very large residual, or unexplained, portion of the wage gap. See BLAU & FERBER, *supra* note 2, at 193; SORENSEN, *supra* note 2, at 3, 46, 135; Joni Hersch, *Male-Female Differences in Hourly Wages: The Role of Human Capital, Working Conditions, and Housework*, 44 INDUS. & LAB. REL. REV. 746 (1991); Hersch & Stratton, *supra* note 25.

important points: that quite severe discrimination can result even when individuals and firms behave rationally, without discriminatory intent, and that focusing on the gender gap as the measure of equality may be the wrong thing to do.

A. *Preliminary Assumptions*

I begin by assuming that individuals and firms are rational. Rational individuals maximize their utility subject to constraints on resources, including their time.²⁷ I assume that rational firms maximize their profits. This is somewhat narrower than the assumption of rationality alone, because it further assumes that markets are perfectly competitive. Thus, any firm pursuing interests other than profit maximization will be priced out of the market.²⁸ This assumption excludes from consideration the possibility that firms are run by misogynistic agents. It also disregards arguments about tastes for discrimination, socialization, and the like.²⁹ The firm described in this model is concerned only with its bottom line, setting marginal costs equal to marginal benefits. These assumptions, no doubt overstated, allow me to present the case against intervention in its most favorable light. Even in this light, we shall see that the case fails.

B. *Phase I: Rational Discrimination*

I begin by looking at the labor supply market in a hypothetical initial period. Because the model is stylized, it need not be situated in any precise historical setting, or given a definite duration. The first issue concerns each worker's marginal productivity, or marginal revenue product. This is the value that an employee is adding to a firm at any point in time. I assume initially that men and women have the same basic marginal productivity function. There is empirical evidence to support this assumption, and at this point in the

27. See Gary Becker, *Irrational Behavior and Economic Theory*, 70 J. POL. ECON. 1, 1 (1962) ("[N]ow everyone more or less agrees that rational behavior simply implies consistent maximization of a well-ordered function, such as a utility or profit function."); John Broome, *Comment: Should a Rational Agent Maximize Expected Utility?*, in *THE LIMITS OF RATIONALITY* 132 (Karen S. Cook & Margaret Levi eds., 1990). See generally Amartya K. Sen, *Rational Fools: A Critique of the Behavioral Foundations of Economic Theory*, 6 PHIL. & PUB. AFF. 317, 322-26 (1977) (discussing and criticizing economic notions of rationality).

28. This is a central point in the reasoning of both Epstein and Posner. See EPSTEIN, *supra* note 3; Posner, *Economic Analysis*, *supra* note 3; Posner, *Efficiency and Efficacy*, *supra* note 3. Posner and Epstein each draw heavily on GARY BECKER, *THE ECONOMICS OF DISCRIMINATION* (2d ed. 1971).

29. For a general discussion of such theories, complete with extensive bibliographic citations, see J. Rubery, *Women's Wages*, in 4 *THE NEW PALGRAVE: A DICTIONARY OF ECONOMICS* 929 (John Eatwell et al. eds., 1987). Technically, my assumptions do not rule out tastes for discrimination, a concept pioneered in BECKER, *supra* note 28, because it is possible that some *claimants* on the firm are discriminatory—that, for example, consumers would pay more for a good produced by an all-male firm, or men would accept lower wages to work in an all-male setting. Examples are discussed in EPSTEIN, *supra* note 3, at 47-58. These effects will not be driven out by the market, but they will, instead, be definitionally efficient.

story I am abstracting from any endogenous human capital decisions.³⁰ Following Kenneth Arrow (and many others), I define discrimination in the marketplace as taking into account factors other than productivity.³¹

I measure discrimination relative to persons having the same productivity. I designate the marginal productivity function as $\beta F(t)$, where $F(t)$ is marginal productivity, a function of time on the particular job, t , and where β represents an intensity parameter, or "multiplier," for weighing marginal productivity.³² I make the typical economic assumptions on $F(t)$, namely that it is monotonically increasing and concave.³³ In plainer words, these assumptions mean that an employee's marginal productivity increases as she spends more time on the job, but at a diminishing rate.

A helpful example is the recent law school graduate beginning her first job in a large, private sector law firm. Initially, she must learn the ropes—not only the substantive body of law in her particular practice area, but also the more formal matters such as how to research, what partners and clients expect of her, and where to go for assistance. During this initial training period, her billing rate is low, and the number of her billed hours may also be low, as partners write off her time.³⁴ As one would expect, all of this changes as she grows into the job. Graphically, the marginal productivity function looks like that depicted in Figure 1.

The vertical axis represents the marginal benefit the employee is providing to the firm at any moment in time. The horizontal axis represents the employee's length of tenure on the particular job. The shape of the curve reflects the fact that an employee's value to her firm is initially quite low—although it need not, and will not generally, be zero³⁵—and then rises rapidly

30. See BLAU & FERBER, *supra* note 2, at 191-95; Donohue, *Economic Perspective*, *supra* note 4, at 1340 n.11; David Neumark, *Employers' Discriminatory Behavior and the Estimation of Wage Discrimination*, 23 J. HUM. RESOURCES 279 (1988).

31. Kenneth J. Arrow, *The Theory of Discrimination*, in DISCRIMINATION IN LABOR MARKETS 3 (Orley Ashenfelter & Albert Rees eds., 1973). See generally Mark Kelman, *Concepts of Discrimination in "General Ability" Job Testing*, 104 HARV. L. REV. 1157, 1164-1203 (1991) (discussing various definitions of labor market discrimination, including, as what he calls static discrimination, the consideration of factors other than marginal productivity).

32. In other words, β can increase or decrease an individual's marginal productivity, depending on whether it is greater than or less than 1. The reason for using β is to take account of the effects of such factors as education and labor market persistence. See *infra* Section III.C.3.

33. See BLAU & FERBER, *supra* note 2, at 146-49, 164-70; see also Robert Topel, *Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority*, 99 J. POL. ECON. 145, 172 (1991) ("[T]he theory of specific human capital [predicts that] compensation rises with job tenure or seniority."). Formally, these assumptions mean that $F'(t) > 0$; $F''(t) < 0$. As I note later, however, nothing really important turns on the exact shape of the marginal productivity curve. The key fact is that there is a period of overpayment that makes tenure critical to the employer's calculations.

34. This "write-off" refers to the practice of partners in law firms not billing their clients for what they—or the client—deem to be excessive time spent on a project.

35. This initial value could be negative, if the costs of training outweighed the benefits provided by the worker in her early period on the job, or positive, if there is an autonomous positive marginal productivity at time 0, as would almost certainly be the case for a later job. However, nothing turns on the absolute magnitude of the figures involved; it is the general shape of the curve that concerns us.

FIGURE 1. *Marginal Productivity Function*

as she learns the job. Over the employee's tenure, her marginal productivity continues to increase, but at a rather slow rate.³⁶ The picture thus represents the familiar learning curve, and is an illustration of the general principle of diminishing marginal returns.

Neoclassical economic theory would hold that Figure 1, representing the employee's marginal benefit to the firm, would also become the employee's wage schedule. In equation form, we would have

$$w = \beta f(t) \quad (1)$$

This equation reflects the general rule of economic theory that sets marginal benefits equal to marginal costs for profit-maximizing firms in a competitive setting. If the firm were to pay the employee more than her marginal contribution to the firm's product, it would lose money. If, in contrast, the firm were to pay the employee less than her marginal product, another firm could lure the employee away with a higher, but still profitable, salary. The offsetting nature of these two forces drives salary structure to an equilibrium where marginal benefits equal costs.

In this first-best setting,³⁷ every employee would be paid her marginal productivity, and we could leave it up to individuals to take actions affecting

36. Over the ranges we are exploring in this model, an employee's value will never decrease with tenure.

37. Throughout the Article, I follow economic convention by using the term "first-best" to refer to an idealized world, without market failures or transaction costs. The real world is invariably second-best.

their productivity. There would be no discrimination. The balance of this Section explores various factors that prevent this situation from coming into existence.

1. *Imperfect Contracts*

In the real world, the transaction costs that neoclassical economic theory abstracts away make the marginal cost/benefit pricing of first-best theory unobtainable. These include, first, the costs of hiring and firing employees—the costs of search and related matters.³⁸ A firm must pay these expenses out of pocket and needs to recoup them over the tenure of the employee. The same is true of any employee-specific investments in human capital that the firm finances.

Second are the costs associated with contracting in the real world. Theory predicts that wage schedules will look like the marginal productivity function of Figure and Equation 1. In practice, however, most market wage structures do not look like that at all. Instead, market wage structures are typically linear.³⁹ The employee is given an initial salary, and, if she persists in the firm, periodic raises. Mathematically, such a wage structure looks something like

$$w = w_i + rt \quad (2)$$

where w is the employee's salary at any point in her tenure with the firm, w_i is her starting salary, and r is the rate of salary increase.

There are several reasons why real-world wage structures are linear. First, it is difficult to design contracts to pay employees based on their marginal productivity at any point in time. In a wide range of employment settings, it may be impossible to measure an employee's marginal productivity. This is especially true where output is a joint product, because determining any individual employee's contribution is very difficult.⁴⁰ Even if an employee's marginal productivity could be monitored and measured, the costs of doing so, combined with the costs of the contracting process itself, would often make such arrangements infeasible. Another set of reasons relates to the preferences,

38. See generally Susan Hanson & Geraldine Pratt, *Job Search and the Occupational Segregation of Women*, 81 ANNALS ASS'N AM. GEOGRAPHERS 229 (1991) (describing differences in ways men and women search for and obtain jobs).

39. Technically, r is unlikely to be a continuous variable; raises are apt to take place periodically, and the pictures in Figures 2 and 3 will involve step functions, not lines. Such contracts are nonetheless called "linear" by convention. See Masahiko Aoki, *Linear Wage Contracts vs. the Spot Market in Their Risk-Bearing Functions*, 30 ECON. STUD. Q. 97 (1979); Donald P. Schwab & Craig A. Olson, *Merit Pay Practices: Implications for Pay-Performance Relationships*, 43 INDUS. & LAB. REL. REV. 237-S (1990) (discussing how merit pay influences the relationship between pay and employee performance).

40. See, e.g., SEN, *supra* note 6, at 119; Stewart Schwab, *Is Statistical Discrimination Efficient?*, 76 AM. ECON. REV. 228, 229 (1986); Amartya Sen, *The Moral Standing of the Market*, 2 SOC. PHIL. & POL'Y 1, 15 (1985).

and risk aversion, of the typical employee. Knowing that a firm can generally terminate her at will,⁴¹ the typical employee will be reluctant to agree to an extended probationary period in which her pay is low, or possibly even negative, to reflect her initially low productivity and to help pay back the search costs.⁴² Instead, employees will want to be assured of a steady income flow over time. Such real-world factors push us almost immediately out of the first-best world of competitive market theory.

Recall the example of the young associate. It might actually be feasible to pay the associate her marginal productivity; the joint-product problem is not severe. The firm could simply give her the value of her billed hours, less some reasonable charge for overhead and training. Easier still, the firm could charge her a fee for its services, and leave the rest to her. But agreeing on the relevant charges and on the procedure for discounting billed time would be cumbersome, at best. Perhaps more to the point, rare is the recent law school graduate who would agree, in the presence of any feasible option, to a contractual arrangement that might pay minimum or even negative wages for six months or so until she learned the job. (Indeed, minimum wage laws by themselves might prevent the firm from paying the associate her marginal productivity.) The very existence of law firms reflects a type of vertical integration of factors of production that has long been associated with the presence of transaction costs in the economy.⁴³ The young associate's salary will move only gradually upwards, even though her marginal productivity is likely to change quite dramatically during her first two or three years in private practice. Salary structure does not reflect marginal productivity very closely in such cases.⁴⁴

41. See EPSTEIN, *supra* note 3, at 147-58. But cf. PAUL C. WEILER, GOVERNING THE WORKPLACE: THE FUTURE OF LABOR AND EMPLOYMENT LAW 48-104 (1990) (discussing modern abrogation of terminable-at-will doctrine).

42. Ronald Coase made this point in his famous 1937 article, *The Nature of the Firm*, reprinted in RONALD H. COASE, THE FIRM, THE MARKET, AND THE LAW 33, 48-51 (1988) (quoting and discussing FRANK KNIGHT, RISK, UNCERTAINTY AND PROFIT (1933)).

43. COASE, *supra* note 42; OLIVER E. WILLIAMSON, THE ECONOMIC INSTITUTIONS OF CAPITALISM (1985). Law firms reflect vertical integration in the sense that individual actors—autonomous agents in neoclassical economic theory—are brought together into a single hierarchy. Thus, secretaries, paralegals, associates and partners (and subclasses of each) are unified into a single firm through a complex web of contracts.

44. I wish to add a few comments for those readers who might insist that real-world wage contracts are far closer to marginal cost pricing than the text indicates. I believe that there are good reasons to be skeptical of this claim, but, in any event, I mean to emphasize that all of the market failures considered in this Part are of a piece. The only relevant points are that (i) in the real world, tenure matters, (ii) firms estimate expected tenure, and (iii) small initial discriminations can lead to significant cycles of discrimination. I hope that the reader who has doubts about any one part of the story can nonetheless see her way through to the whole.

2. *Incomplete and Asymmetric Information*

The first problem, that of the infeasibility of designing theoretically perfect employment contracts, leads almost immediately to a second, that of imperfect or incomplete information.⁴⁵ Under the linear wage structure discussed above, the employer's ex ante prediction of an employee's likely tenure with the firm becomes critical. This was not so under the first-best world of ideal theory, where the firm had no search costs and was breaking even with its employees at all times. Under the linear wage structure, the firm must overpay the employee during an initial phase, and depend on a subsequent period of underpayment to recoup its initial losses.⁴⁶ The juxtaposition of the marginal productivity function and the linear wage schedule looks like Figure 2.

FIGURE 2. *Marginal Productivity and Linear Wage Schedule*

We can now see, graphically, that the wage schedule is above the productivity of the worker until a certain threshold point, labelled *a* in Figure

45. For a discussion of the technical meanings of various types of information—such as incomplete and imperfect information—see ERIC RASMUSEN, *GAMES AND INFORMATION: AN INTRODUCTION TO GAME THEORY* 51-60 (1989). I discuss my (somewhat idiosyncratic) use of the distinction in the text below.

46. By the same logic developed in the text, firms will have to design contracts that assure that employees will persist during this second, underpaid, period. This goal might be accomplished, for example, through seniority-based pensions plans or the promise of an eventual period of *overpayment*, such as on making partner or retiring. The important differences between piece-work and deferred compensation arrangements are discussed in Goldin, *supra* note 17, at 4-8. For another discussion of these dynamics, see Jeremy I. Bulow & Lawrence H. Summers, *A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment*, 4 J. LAB. ECON. 376 (1986).

2, is met. The employer's ex ante concern extends beyond point a , however, for the firm must recoup its losses incurred up to that point. The rational firm wants some assurance that a potential employee will persist at least until time t^* , when the salary and productivity functions will have reached some point, labelled b in Figure 2, such that the initial overpayment is recouped by the later underpayment.⁴⁷

The critical problem here is ascertaining how long an employee is likely to stay with the firm. This is a problem of both incomplete and asymmetric information. The employee herself does not know with any certainty her likely tenure, and the firm is at an even greater disadvantage, because it cannot trust the employee to be fully honest. Later, I consider the type of signals a potential employee might send regarding her likely persistence with the firm. For now, in the hypothetical initial period, I note simply that the firm has to guess. It is almost certain that the firm will guess that women, and especially married women, will have an average tenure with the firm shorter than that of men. The reason, of course, is that firms believe that these women will have children and leave the work force, at least temporarily. There is abundant evidence that firms think this way,⁴⁸ and the facts—although we are now well beyond any hypothetical first period—support the general conclusion: women have higher turnover rates, and less persistence.⁴⁹ Indeed, during significant periods of American history state laws required married women to leave their jobs,⁵⁰ so that the only guessing an employer had to do was whether or not a female employee would marry. Even today, with declining marriage and fertility rates, 85% of all women bear children.⁵¹ Evidence supports the fact, consistent with this model, that women who were once married—whether they are now married, divorced or widowed—have a sizeable gender gap, whereas women who never married do not.⁵²

47. Figure 2 is meant to be drawn so that the area below the wage line but above the marginal productivity curve before point a is equal to the area above the line but below productivity between points a and b .

48. See, e.g., GOLDIN, *supra* note 2, at 176 ("The sequel to the marriage bar was the pregnancy bar. All in all, the best female employee was, in the words of a Sears, Roebuck, and Co. officer 'a married woman with a mortgage on her house and her children partially raised.'"); SCHWARTZ, *supra* note 7, at 9-26, 40-50; Bulow & Summers, *supra* note 46, at 399-401; Goldin, *supra* note 17, at 2-3; Light & Ureta, *supra* note 17.

49. BLAU & FERBER, *supra* note 2, at 78-80, 150-54; SORENSEN, *supra* note 2, at 105-28; Joyce P. Jacobsen & Lawrence M. Levin, *The Effects of Intermittent Labor Force Attachment on Female Earnings*, Paper Presented at the American Economic Association Conference (Jan. 1992); Maume, *supra* note 17, at 500-05. But see BERGMANN, *supra* note 7, at 113 (discussing studies showing that turnover rates are similar for men and women); Donohue, *Economic Perspective*, *supra* note 4, at 1367 n.58 (suggesting that women no longer have higher turnover rates than do men). If this latter claim is true, this supports the general idea that the gender gap is narrowing because women in the work force are acting as men typically have.

50. See GOLDIN, *supra* note 2, at 159-79.

51. See LINDEN, *supra* note 2, at 4 (noting that 82.3% of women ages 35-39 in 1990 were mothers); Felice N. Schwartz, *Women as a Business Imperative*, HARV. BUS. REV., Mar.-Apr. 1992, at 105, 111.

52. See Fischel & Lazear, *supra* note 3, at 899 ("Women who have never been married appear to have complete wage parity with men who have never been married. Conversely, women who are currently

Within the first-best, optimal contract setting, the potential tenure of the employee did not matter, because she would at all times be paid her marginal contribution to the firm. Imagine a piece-work, or commission-based compensation scheme—expected tenure is irrelevant.⁵³ In contrast, under a typical real-world wage structure, expected tenure is crucial. If we simplify the analysis by holding the rate of salary increase constant, the initial wage, w_i , becomes a strictly increasing function of projected tenure.⁵⁴ If the firm thinks that it can count on the employee to stay longer, it can afford to pay her a higher salary up front, or it can invest more in her training, because it has greater confidence in its ability to recoup these costs later. The firm will have a wage structure that varies with the employees' persistence profiles.

Because the firm cannot know with any certainty which married women of childbearing age will go on to have children and take maternity leave or quit their jobs altogether, it bases its hiring decisions on the average tendencies. This is efficient in the sense that the firm is making a rational, cost-benefit decision—the costs of getting better information do not equal the benefits of such information *to the firm*—but the general optimizing condition that marginal benefits equal costs in production has already been abandoned. It is not just that some women become pregnant, and no men do, or that nature somehow dictates that women should take time off to parent while men should not.⁵⁵ If we were in a first-best setting, a person would always be getting paid her marginal productivity, and we could leave it up to individuals to take actions that affected their wage levels, one way or another. Recall again the piece-work model. But in the real world, women who plan on having uninterrupted careers—or are even uncertain about their career paths—suffer from the inability of the market to screen out those who will in fact have short tenures.

married earn less than half of what men who are currently married earn.”); Hill, *supra* note 10.

53. See, e.g., Goldin, *supra* note 17.

54. A simple way to see this point is to abstract to the pure case where r in textual equation (2) is zero—the employee receives no increase in her salary. This might indeed be the limiting case to which a risk-averse employee and a risk-neutral employer would come, but for the employer's desire to provide incentives for the employee to continue improving her productivity. In such a case, equation (2) becomes

$$w = w_i$$

which is a horizontal line, with 0 slope. Now since, by assumption, $f(t)$ is monotonically increasing, its slope never reaches 0. This means that, for each marginal interval of time for which

$$w < f(t)$$

the employer is making a profit on the employee's labor. But since the competitive market assumption rules out a *net* positive return, this positive period must be offset by a negative return earlier on in the employee's tenure. This negative return can only be provided by a higher initial salary. The discussion at note 69 *infra* relaxes the assumption that r is constant, and we shall see that gender-specific wage tenure profiles may aggravate the self-selection property to be discussed below.

55. For examples of this argument, see EPSTEIN, *supra* note 3, at 269-74; Posner, *Economic Analysis*, *supra* note 3, at 1314-15. See generally BLAU & FERBER, *supra* note 2, at 14-18 (discussing sociobiological arguments in gender analysis).

This failure of firms to screen out those with shorter tenures is termed "statistical discrimination." It results from firms' relying on probability assessments rather than individual circumstances and has flourished because it is more cost-effective than efforts to acquire better information.⁵⁶ Statistical discrimination, at least as I am using the term, refers not only to decisions not to hire certain groups, but also to decisions regarding salary level and the allocation of employees to jobs within the firm. Statistical discrimination might lead to lower pay levels for women, or to steeper wage-tenure profiles, or to a concentration in nonmanagement positions, and so on.⁵⁷ If these practices are constrained by antidiscrimination laws, such as those requiring equal pay, greater (private) inefficiencies might result.⁵⁸

Statistical discrimination is sometimes confused with, or given the label of, "efficient" discrimination.⁵⁹ But this type of discrimination is only conceivably efficient in the normatively questionable Kaldor-Hicks, or wealth-maximizing, sense.⁶⁰ Perhaps more importantly, if the discrimination works against a class with a higher wage elasticity—and there are very good reasons to believe that it does in the case of gender discrimination—it will flunk even this test.⁶¹ The move may be rational or cost-minimizing *to the firm*, but this does not mean that it is efficient in a public sense. The distinction between private and public efficiency is critical; the private efficiency or rationality of

56. See GOLDIN, *supra* note 2, at 89, 214; Arrow, *supra* note 31; Edmund S. Phelps, *The Statistical Theory of Racism and Sexism*, 62 AM. ECON. REV. 659 (1972). For critical discussions, see Becker, *supra* note 27; Kelman, *supra* note 31.

57. See Bulow & Summers, *supra* note 46, at 399-400; Joanne Salop & Steven Salop, *Self-Selection and Turnover in the Labor Market*, 90 Q.J. ECON. 619 (1976); Joni Hersch & Patricia Reagan, *Efficient Gender-Specific Wage-Tenure Profiles* (1993) (unpublished manuscript, on file with author).

58. This argument is at least implicit in Posner and Epstein's work. See EPSTEIN, *supra* note 3; Posner, *Economic Analysis*, *supra* note 3; Posner, *Efficiency and Efficacy*, *supra* note 3.

59. See, e.g., Becker, *supra* note 27, at 936-37. Becker goes on to note that statistical discrimination, even if accurate, may not be efficient. *Id.* at 937. Epstein concedes that statistical discrimination involves inefficiency, but believes that the possible cures are worse than the harms. EPSTEIN, *supra* note 3, at 37-41.

60. See *infra* Section IV.A.

61. Stewart Schwab has suggested that statistical discrimination helps one class and hurts another, so that, in general, it is not possible to predict the efficiency gains a priori. Schwab, *supra* note 40, at 229. Schwab goes on to note that, if the harmed class is more elastic than the benefitted class, the discrimination will be inefficient. *Id.* at 232-33. (This point assumes that the size of the classes is equal.) I pick up this point in Part V, once we have considered taxes and other market failures, but we can already note that the higher elasticity of the discriminated-against class is highly likely within our model, in part because the trait being screened for is labor market commitment. See also Stephen Coate & Glenn Loury, *Will Affirmative Action Policies Eliminate Negative Stereotypes?*, AM. ECON. REV. (forthcoming 1993) (modelling self-fulfilling prophecies); Donohue, *Economic Perspective*, *supra* note 4; Mayer G. Freed & Daniel D. Polsby, *Privacy, Efficiency, and The Equality of Men and Women: A Revisionist View of Sex Discrimination in Employment*, 1981 AM. B. FOUND. RES. J. 585, 633-35 (discussing privately efficient statistical discrimination that leads to self-fulfilling prophecies, using the example of female job persistence); Shelly J. Lundberg & Richard Startz, *Private Discrimination and Social Intervention in Competitive Labor Markets*, 73 AM. ECON. REV. 340, 347 (1983) ("It is not an appropriate response, even for those of us who generally believe in the efficacy of private markets, to dismiss discrimination as something 'the market will handle.'"). For a different argument on the inefficiency of statistical discrimination, see Bulow & Summers, *supra* note 46, at 383-90, 398-401 (suggesting that firms are constrained from offering low wages by a shirking problem, and that therefore inefficiently high wages may persist in real-world markets).

firm conduct has little normative significance in a world of market failures. For example, a firm that continues to emit polluted smoke in the absence of laws compelling it to pay the associated costs is acting rationally, or efficiently, in a private sense. Few would argue, at least on first-best grounds, that this is an efficient outcome, or that we should respond by doing nothing.

The same type of logic applies to the sort of informational failures at issue with gender and the work force. In and of themselves, these conditions prevent the rationality of firm actions from having decisive normative force. Ideal, first-best market theory rests on the simple idea that the rational self-interested action of all agents can, with minimal state action, lead to the best of all possible worlds. Market failures undermine this simple moral equivalence. Where transaction costs vitiate the strong efficiency claims of free markets, the fact that firms act rationally in the face of such costs cannot be dispositive of normative argument. In Part IV, I consider the more plausible, second-order version of this argument: Market failures, including informational ones, theoretically justify interventions, but the costs of actual interventions typically exceed their benefits. This calls for a different type of response, balancing real-world costs and effects.

I want to deviate from the main discussion for a moment to comment on Epstein's analysis of statistical discrimination in his recent book, *Forbidden Grounds*.⁶² While conceding both the existence and inefficiency of statistical discrimination, Epstein finds the theory of limited use in explaining real-world outcomes, or in justifying market intervention.⁶³ He argues that parties have suitable incentives and opportunities, within a free market structure, to overcome the informational problems that lead to statistical discrimination. Epstein does make what I find an interesting assertion—that antidiscrimination laws may make the problem of statistical discrimination *worse*, by prohibiting employers from obtaining more and better information⁶⁴—and this becomes part of the case against traditional regulatory solutions. But Epstein's basic analysis in minimizing the role of statistical discrimination suffers from three fatal flaws.

First, Epstein fails to distinguish between imperfect and incomplete information. Imperfect information is information that is knowable but not in fact known to a given party. It can be cured by a suitable investment of resources to make the knowable known.⁶⁵ Incomplete information is information that is neither known nor knowable; in such a case, no investment of resources can cure the problem. One's ability, for example, is a question of

62. EPSTEIN, *supra* note 3, at 33-41.

63. *Id.* at 40-41.

64. *Id.* at 40. *But cf.* Freed & Polsby, *supra* note 61; Kelman, *supra* note 31; Lundberg & Startz, *supra* note 61.

65. For a discussion of the terms, see RASMUSEN, *supra* note 45, at 51-60. I use the term "imperfect" somewhat differently—and less technically—than Rasmusen does.

imperfect information. Ability is knowable, but not necessarily known to all relevant parties. In contrast, the weather or general price level at some time in the distant future is a case of incompleteness: no one really knows. The issue in the model of gender discrimination is precisely this type of incomplete information. No one can know how long an employee will stay on the job, and no investment of time or money will clearly improve one's predictive powers in this regard.⁶⁶ There is no perfect test for persistence. In such a setting, a large residual is left for guessing, no matter how educated such guessing might become.

Second, while Epstein does take the next step in informational economics—the move to signaling—he curiously ignores its costs. Signaling refers to actions that an informed party takes to convey information to another party. Offering a guarantee on a product, for example, conveys the guarantor's confidence in its quality. I discuss below the types of signals, most importantly education, that women might use to demonstrate their likely persistence. Signaling, if rationally engaged in, may improve market outcomes, but it has a cost.⁶⁷ Especially where only one class of actors—here, women—are put to the burden of signaling, and where the signaling involves significant real expense, one cannot gloss over the normative propriety of such costs. It is as if the market were telling women that they could get high quality jobs, just like men, but first they had to pay \$50,000. The fact that the barrier *could* be overcome, so that women could get the jobs, hardly justifies ignoring the existence of the barrier altogether.

Third, Epstein's analysis at this juncture is notably static. The residual left for statistical discrimination, after the parties take all rational actions in the face of the underlying informational deficiencies, may indeed be small. I do not think that it is, because of the two factors already noted. Conceding this point for the sake of argument, however, the *dynamic* consequences of even small initial discriminations, once a rich social and institutional fabric is added, can be quite dramatic.⁶⁸ This will become one of the central themes throughout the remainder of this Part. For all of these reasons, the consequences of statistical discrimination in a dynamic setting with incomplete information and high signaling costs can be quite large.

Returning to the model, we can simplify for now to the case where most women are viewed as potential mothers; those women who are not so viewed

66. See Posner, *Economic Analysis*, *supra* note 3, at 1320 ("[W]hile it would not be efficient to ascertain individual qualifications, it would be *possible* to do so—that is, the cost would not be infinite. But in the case of women, the cost *would* be infinite.").

67. *Id.* at 205-17; see also A. MICHAEL SPENCE, MARKET SIGNALING 176 (1974) ("The productivity gains attributable to the information content of market signals may or may not justify the resource cost. Certainly one cannot assume that signaling is more efficient than no-signaling.").

68. See GARY BECKER, A TREATISE ON THE FAMILY 4 (enlarged ed. 1991) ("[T]he message is that even small amounts of market discrimination against women or small biological differences between men and women can cause huge differences in the activities of husbands and wives.").

FIGURE 3. *Male and Female Linear Wage Schedules*

are treated, at least currently, like men. The first phase ends with a two-track labor market. Lower initial wages are offered to women because they are expected to have shorter tenures. The picture looks like Figure 3, where the higher wage schedule is that for men, for whom the employer predicts a tenure of t^* , and the lower wage schedule is for females, whom the employer predicts will work for t^{**} .

We should be clear on the nature of the "discrimination" at this point in the story, all of which has been due to the rational action of firms. Women who in fact work until t^{**} —just like men who work until t^* —will be paid exactly their marginal product, in the aggregate, over that period. Women who work less than t^{**} —and men who work less than t^* —will be overpaid. Women who plan to and in fact do work beyond t^{**} , say to t^* , are underpaid, relative both to women who work less and to men who work just as much. If the typical employer's assumption is that men will work indefinitely, whereas women will not, this latter class is quite large—there are many more women than men who "lose" because of statistical discrimination, even if many of the "winners" are also women. What may be more important than this aspect of discrimination, as the subsequent discussion aims to establish, is the *ex ante* setting of the wage structure. This institutional rigidity turns out to have very important cyclical effects, as rational individuals make their life and career decisions on the basis of an offered wage structure that features lower pay for women.

In any event, this first phase has highlighted three points. In the real world, as compared to the ideal world of first-best theory: (i) tenure matters, (ii) firms must estimate expected tenure, and (iii) such estimations are likely to disfavor women, especially women contemplating longer tenures. Labor economists, among others, have developed much more sophisticated models, building in specific wage-tenure profiles, endogenous human capital decisions, incentives to shirk, and so on.⁶⁹ But the two key factors driving all of these analyses are that estimated persistence is important and that women are thought to be less persistent. History is also relevant, because, as I demonstrate below, these conditions, once in place, constrain the course of future market developments.⁷⁰

C. Phase II: Self-Fulfilling Prophecy

As the analysis outlined above suggests, a variety of forces interact to create a self-perpetuating cycle, as many commentators have noted.⁷¹ The result is a type of chicken-and-egg problem of which came first: women's preferences over work patterns or employers' expectations of lesser female

69. For example, the gender-specific wage-tenure contract, which might be quite rational under the circumstances, has steeper wage-tenure profiles for women; this relaxes the assumption of constant r used in the text. Thus, the wage-tenure profile for men and women might look something like Figure 4:

FIGURE 4. Wage Tenure Profile with Steeper Female Schedule

Among other features, such a wage-tenure profile might induce even more self-selection among women, and influence women in the work force to stay there all the more; these self-perpetuating phenomena are discussed in Section III.D, *infra*. See Hersch & Reagan, *supra* note 57, and sources cited therein; Salop & Salop, *supra* note 57. Of course, precisely this type of gender-specific wage schedule might be barred by the equal pay aspects of antidiscrimination laws. Another problem is that the lower wages called for in the model may be constrained by the need to eliminate shirking. Bulow & Summers, *supra* note 46, at 383. For other labor economic perspectives, see Lazear & Rosen, *supra* note 14; Topel, *supra* note 33. For an argument as to why women might accept less steep wage-tenure profiles than men, due to their greater "tastes for cooperation," see Rose, *supra* note 11, at 433-34.

70. Note that details such as the precise shape of marginal productivity curves or linear wage structures are not critical to these basic points.

71. See, e.g., GOLDIN, *supra* note 2, at 214; HOCHSCHILD, *supra* note 1, at 254; Becker, *supra* note 4; Coate & Loury, *supra* note 61, at 2; Freed & Polsby, *supra* note 61, at 634-35; Reuben Gronau, *Sex Related Wage Differentials and Women's Interrupted Labor Careers—The Chicken or the Egg*, 6 J. LAB. ECON. 277, 294 (1988); see also Rose, *supra* note 11, at 443-54 (discussing cycles against women, perhaps based on self-fulfilling prophecies).

persistence.⁷² Women get paid less and therefore want to work less. This will be true as long as the wage elasticity of labor supply is positive—meaning that a decrease in wages will lead to a decrease in labor supplied.⁷³ This is in fact generally true for all workers, and is especially true for married women, who typically have quite high labor supply elasticities.⁷⁴

Indeed, the analysis of Section III.B can explain even more: men gravitate towards professions where tenure is especially important, and reap the benefits of this concentration; women flood the markets where tenure is less important, depressing even further the wage levels in these presumably lesser-skilled positions.⁷⁵ Firms, constrained by law or custom to offer similar wages to all employees, might avoid hiring women altogether: Posner and Epstein may be right that there is a certain counterproductive nature to aspects of the antidiscrimination laws.⁷⁶ What I want to stress is that market failures have already been a major part of the story in this first phase. Search costs and imperfect contracts created transaction costs. Incomplete, asymmetric information, which led first to a need to estimate persistence and then to an assumption of lesser female persistence, compounded the problem. Whatever else may have been going on, these factors played some role. We were never in a first-best setting where we could make strong efficiency claims.

Once basic discrimination appears in the wage structure, there are reasons why it might perpetuate itself. Being paid less, women are in fact more likely to drop out of the work force for stretches, confirming their employers' prior expectations, again assuming positive labor supply elasticity. Facing a lower *ex ante* wage schedule, women might also invest less in education and training, and therefore be more likely to become the marginal earner in two-earner families, although later I explore how lower wages may have precisely the opposite effect on some women's educational decisions.

This Section adds three more factors to the dynamic story. First are taxes, that encourage the marginalization of one spouse in a marriage. Epstein, Posner, and most others staking out a non-interventionist posture generally

72. See GOLDIN, *supra* note 2, at 184; Cain, *supra* note 26, at 118; Melvin W. Reder, *Comment*, in DISCRIMINATION IN LABOR MARKETS, *supra* note 31, at 34-35.

73. See *infra* Section V.A (discussing elasticities).

74. See Janet C. Hunt et al., *Taxation and the Wife's Use of Time*, 34 INDUS. & LAB. REL. REV. 426, 432 (1981); Jane H. Leuthold, *Home Production and the Tax System*, 3 J. ECON. PSYCH. 145 (1983); Harvey S. Rosen, *Tax Illusion and the Labor Supply of Married Women*, 58 REV. ECON. & STAT. 167 (1976); Robert K. Triest, *The Effect of Income Taxation on Labor Supply in the United States*, 25 J. HUM. RESOURCES 491, 510-13 (1990). See generally Jerry A. Hausman, *Taxes and Labor Supply*, in HANDBOOK OF PUBLIC ECONOMICS 213, 247-49 (Alan J. Auerbach & Martin Feldstein eds., 1985) (discussing evidence of high labor supply elasticity of wives).

75. Presumably, the reverse cannot be as true. Men cannot flood the higher-skilled, longer-tenure positions, because there are constraints on their skill levels, or other preference-based choices, that keep them from making this move. Otherwise no men would presently be in low-skilled jobs.

76. See EPSTEIN, *supra* note 3, at 242-66; Coate & Loury, *supra* note 61, at 2-3; Fischel & Lazear, *supra* note 3, at 913-15; Posner, *Efficiency and Efficacy*, *supra* note 3, at 521.

leave taxes out of their analyses,⁷⁷ but I believe that taxes are crucial, both in themselves and in their connections with the informational failures chronicled above. Second are incomplete markets, which offer few and relatively unattractive options in part-time and flexible time careers. Together, the tax system and incomplete markets reinforce the problems already discussed, further demonstrating how market failure, rather than market efficiency, is central to much of the descriptive story of women in the labor force. Finally, I discuss human capital decisions.

1. *Taxes*

I have argued in a related article that the tax system has strong behavioral biases that are far more dramatic than commonly thought.⁷⁸ I do not want to repeat that analysis here; a summary of the findings should suffice. Five factors in the tax law contribute to a pattern of behavioral incentives. One, tax law aggregates the incomes of husband and wife, placing the marginal worker spouse in an initial tax bracket dictated by the primary worker's wages. Two, the law does not tax imputed income from self-supplied services, giving individuals an incentive to stay home and provide their own home and child-care work. Three, the structure of social security contributions and benefits protects stay-at-home spouses while dramatically penalizing secondary earners. Four, the law provides inadequate relief for mixed business-personal expenses, especially child care, aggravating the biases mentioned above. Five, the rules governing taxation of noncash benefits encourage single-earner spouses to provide health, pension, and related benefits to the entire family.⁷⁹

These factors lead to three distinct behavioral effects. First, the tax laws discourage couples from marrying in the first place, at least in any situation where both spouses might work. Second, and related, the law provides a strong incentive to have only one spouse work. Finally, the law induces all-or-nothing labor decisions by secondary earners—that is, discourages secondary earners from working part-time.⁸⁰ In general, the tax laws, conjoined with real-world wage structures, push women to the margins of their household economies, and strongly influence their choices thereafter. The precise nature of the effects varies with income level. The incentive against marriage is especially strong

77. Epstein leaves taxes out of his analysis altogether. See EPSTEIN, *supra* note 3. Posner leaves them out in most of his work, although he does mention them in *Conservative Feminism*, 1989 U. CHI. LEGAL F. 191 (noting distortion created by failure to tax housewives' imputed earnings).

78. For the lengthier analysis that forms the basis of the summary provided here, see Edward J. McCaffery, *Taxation and the Family: A Fresh Look at Behavioral Gender Biases in the Code*, 40 UCLA L. REV. 983 (1993). Under the new 1993 tax law, the marriage penalty—which, I argued, may be the wrong thing to focus on in any event—will become even more severe. See Chet Currier, *Tax Rules Influence Love, Marriage and Divorce*, L.A. TIMES, Nov. 29, 1993, at D6.

79. McCaffery, *supra* note 78, at 988-1013 and sources cited therein.

80. *Id.* at 1014.

among lower income families; the all-or-nothing effect is especially present among middle income families; and the push towards single-earner households is most marked among upper income households.

These inducements are not minor matters, or mere scholastic quibbles. On the one hand, one might be tempted to see the tax system as simply a proxy for the whole set of laws inherited from a male-dominated world that adds to the informational failures chronicled above to perpetuate a vicious cycle against women.⁸¹ That the tax laws are just one among many villains might indeed be the case; I have no doubt that other bodies of laws have pronounced effects on the gender status of the work force. But the tax system in and of itself is massive, as are its particular behavioral effects. A few numerical examples from my previous article should help drive home these ideas.

Under 1991 tax rates—more recent numbers look even worse for women—a spouse in a lower income household, where the other spouse makes around \$14,000 a year, faces a marginal tax rate of approximately 50% on her *first* dollar of earned income. For a middle income wife whose husband makes \$25,000 a year, a part-time position paying \$8500 a year will net the family around \$152 a month, after taxes and all work-related expenses; in a wide range of actual cases, the wife's part-time work actually *loses* money for the family. For an upper income family where the husband makes \$60,000 a year, the wife would have to earn \$30,000 to net the same amount of after-tax dollars for the family as the husband would by making an additional \$12,000, about one-third as much as his wife.⁸²

These are dramatic numbers. My previous article underscored the importance of these figures in setting tax policy, and my point here is that we cannot ignore them in analyzing existing labor market patterns. All real-world taxes, other than market-correcting Pigouvian taxes,⁸³ result in market failure by driving a wedge between marginal benefits and costs and violating the conditions needed for the neoclassical welfare theorems.⁸⁴ Few would dispute that taxes of the magnitude described above affect behavior.⁸⁵

The present system of taxation plays a more particular role in the gender discrimination story. First, it hastens the push toward identifying household earners as primary and secondary. The large after-tax benefits for being a single-earner family encourage families to identify just one earner to cash in

81. See, e.g., GOLDIN, *supra* note 2, at 186-99 (suggesting that protective legislation in America had the unintended consequence of holding women back); Sylvia A. Law, *Women, Work, Welfare, and the Preservation of Patriarchy*, 131 U. PA. L. REV. 1249, 1252-62 (1983) (discussing gender biases in labor and welfare policies).

82. McCaffery, *supra* note 78, at 1014-29.

83. See A.C. PIGOU, *THE ECONOMICS OF WELFARE* 192-95 (4th ed. 1948); William V. Baumol, *On Taxation and the Control of Externalities*, 62 AM. ECON. REV. 307 (1972).

84. See ROBIN W. BOADWAY & NEIL BRUCE, *WELFARE ECONOMICS* 3 (1984); DAVID M. KREPS, *A COURSE IN MICROECONOMIC THEORY* 289-91 (1990).

85. See *DO TAXES MATTER?* (Joel Slemrod ed., 1990).

on these potential benefits. Because the wife is apt to be the lower earner and hence the secondary worker, she has an initial marginal tax rate that is at least as high as her husband's. She therefore faces a significantly higher average tax rate.⁸⁶ The laws therefore make the wedge between her productivity and compensation in the *paid* market especially high. Even with a labor supply elasticity no greater than her husband's, the wife faces greater pressure to move out of the job market as a result of the higher tax burden. Further, the wife is apt to compare her paid market work, which is reduced by both statistical discrimination and taxes, to unpaid domestic work, which is affected by neither factor. Add to this scenario the facts that entering the labor force makes these previously untaxed household services taxable, that social security works as a pure tax on the secondary earner over her initial range of wages, and that tax laws have given an incentive to have the husband provide noncash or fringe benefits to the whole family, and the skew in incentives becomes clear.

In short, our present tax system rewards and encourages single-earner families. This bias puts pressure on families to identify a primary worker and a secondary or marginal one. The market itself has simultaneously weighed in on this score through rational, statistical discrimination against women, encouraging families to view the wife as the marginal worker by paying her less. Once this bias is in place, tax laws compound the problem by giving further reason for the wife to flee the job market and for the husband to work harder. This outcome confirms employers' initial expectations, and the cycle continues.

2. *Incomplete Markets*

Another deviation from ideal market theory involves the absence of a set of complete markets in labor. Some may see this absence as a market failure in and of itself,⁸⁷ but it is just as easy to see the absence of a full range of opportunities as the result of other, prior market failures. Something is going on, such as the presence of transaction costs or taxes, that is obstructing the conditions of first-best theory.

Consider the general case of a married mother who has just had a child, perhaps her first. We expect that she loves her child and wants to spend time with him. She also faces tax and market-based incentives for staying home and

86. Normally, marginal rates are the chief concern in assessing the distortionary effects of tax. For an example of this, see the discussion of optimal taxation in Section V.A, *infra*. In making her decision to work or not, however, a rational taxpayer would look at the average rate borne over the range of anticipated income.

87. In the rarefied general equilibrium theory of Arrow, Gerard Debreau, and others, a set of complete markets is a prerequisite to the strong findings of the welfare theorems. See KREPS, *supra* note 84, at 193-94; David M. Newberry, *Missing Markets: Consequences and Remedies*, in *THE ECONOMICS OF MISSING MARKETS, INFORMATION, AND GAMES* 211 (Frank Hahn ed., 1989).

providing untaxed nonmarket labor. At the same time, however, she might well desire money income, the maintenance of job market skills, and perhaps some respite from the often overwhelming burden of child care. If we were visiting from Mars, and knew only these facts, we might conclude that our generic new mother would look for part-time work.⁸⁸ Indeed, engaging in the luxury of a priori speculation a bit further, we might conclude that *both* spouses would cut back on their labor force participation to share in the joys and stresses of child-rearing, while each contributed money income and maintained workplace skills. But neither of these patterns develops very often in America. Part-time work in general remains poorly paid and of low prestige. Most of the growth in part-time employment over the last twenty years has been involuntary, typically in situations where the employee desires full-time work but the employer does not offer it.⁸⁹ Relatively few married mothers avail themselves of the part-time employment option, and the two part-time earner household is exceedingly rare.⁹⁰

At first glance, this just seems to be a cultural oddity, particularly since more flexible models of part-time employment are more common in other societies.⁹¹ Deeper analysis, however, reveals that the absence of better part-time options is not only something of an embarrassment for market theory, but also a significant problem in itself. It is an embarrassment because theory would seem to indicate that, given a plausible a priori case for the demand for such positions, a market would develop, and price and other non-price terms would respond accordingly. Of course, if the marginal productivity of part-time labor were somehow lower than that of full-time labor—if for example, individual production functions dictated that individuals were more marginally productive working forty-hour weeks than twenty-hour ones—this difference would explain the gap. The evidence of such effects, however, is inconclusive at best. Putting aside for now questions as to why there is not a more vibrant part-time market—I have implied that taxation is one answer, and there are no doubt many others on the firm, or demand side of the equation—the very

88. We might also want to know certain facts about the relative returns to scale of domestic and labor market activity. In the unlikely event that these returns were steadily increasing, specialization of labor might be efficient. In the more plausible situation where returns were constant but there was a fixed initial cost of work—e.g., in learning the necessary skills to enter—specialization could also be efficient. These and related matters are discussed in BECKER, *supra* note 68.

89. TILLY, *supra* note 21, at 1, 13-20; Susan Diesenhouse, *In a Shaky Economy, Even Professionals Are Temps*, N.Y. TIMES, May 16, 1993, at K5.

90. Of 43,450,000 married-couple families in 1987, only 637,000 (less than 1.5%) fit the two part-time earners pattern; this number no doubt includes many couples without children living at home. U.S. DEP'T. OF COMMERCE, BUREAU OF THE CENSUS, STATISTICAL ABSTRACT OF THE UNITED STATES 459 (Table No. 738) (111th ed. 1991). Tilly notes that women of childbearing age have decreased their participation rate in the part-time market since 1969. TILLY, *supra* note 21, at 1. As of 1989, 67% of the mothers who work had full-time jobs, the same percentage as in 1959. For other views on the role of women in the part-time labor market, see GOLDIN, *supra* note 2, at 183 ("[W]omen who work part-time today earn about as much per hour as do those who work full-time."); HOCHSCHILD, *supra* note 1, at 2.

91. BLAU & FERBER, *supra* note 2, at 317; JOHN D. OWEN, REDUCED WORKING HOURS: CURE FOR UNEMPLOYMENT OR ECONOMIC BURDEN? (1989); Chamallas, *supra* note 21, at 722 n.87.

absence of such opportunities is itself a problem moving us away from the first-best setting.

One of the conditions needed to support the first welfare theorem, which essentially holds that free trading leads to Pareto optima, is the presence of a set of complete markets.⁹² The absence of part-time options means that there are incomplete, or missing, labor markets. Employees do not have a choice along a smooth spectrum of hours, deciding exactly how long they want to work. Instead, most opportunities cluster around forty-hour work weeks, with only low quality part-time ones in the middle, so that employees who want to work less must make second-best choices. Like the absence of a set of complete contingent contracts or the presence of indivisibilities in expenditure, this incompleteness invalidates the claim for efficiency of the market structure as it now exists.⁹³

The absence of viable part-time options, combined with a tax-compounded set of incentives for not pursuing the relatively unattractive part-time options that do exist, means that the typical married mother faces an all-or-nothing labor choice. Many women will resolve this dilemma in favor of not working at all. Forty percent of married mothers with children under the age of six stay home.⁹⁴ Consequently, because of the costs of rearing children, married fathers of young children tend to work harder, to replace the income from the wives' reduction in work. This pattern reinforces the employers' initial probabilistic assessment: married women are poor risks, but married men are attractive ones. Indeed, the hierarchy of wages in the labor market is quite clear: married men are the best paid; followed by never-married men and women, who have virtual wage parity;⁹⁵ with ever-married women bringing up the rear. I should thus add a third and higher line to Figure 3, above, for married men, leaving never-married persons of either sex in the middle.

Before one can write off the general labor market situation as being efficient, we need to consider that the absence of a vibrant part-time market in America has contributed to the choices workers, and especially parents, have made. One might be tempted to argue, or infer, that the very absence of better part-time options is because of individual preferences; rational actors simply do not want to work part-time. Before we can justify that assertion, however, we need to consider that the landscape has already been shaped by transaction costs, incomplete information, and taxation, to which we might add the collective action problems of forming high quality part-time markets.⁹⁶ There are

92. See KREPS, *supra* note 84, at 193-99; HAL R. VARIAN, MICROECONOMIC ANALYSIS 365-66 (3d ed. 1992).

93. KREPS, *supra* note 84, at 193-95; Newberry, *supra* note 87, at 212-14.

94. See STATISTICAL ABSTRACT, *supra* note 15, at 388 (Table Nos. 620, 621).

95. Fischel & Lazear, *supra* note 3, at 899.

96. See Rose, *supra* note 11, at 439-40 (discussing collective action problem in challenging stereotypes or conventions and citing, inter alia, Robert Sugden, *Contractarianism and Norms*, 100 ETHICS 768, 779-82 (1990)).

barriers to creating part-time work on both the firm and the individual side that have no place in first-best theory.⁹⁷ The proof of the market failure is both easy and intuitive: if there is even a single person, anywhere, who would willingly take a high quality part-time position, paying her no more than her marginal productivity, if such a position were available—and most of us know such a person—then the absence of such an opportunity is a market failure.

3. *Human Capital Decisions*

I want to bring human capital decisions into the story at this point. Recall β , the parameter on each worker's marginal productivity function from Equation (1). I now inject more realism into the model by making β itself a function of two variables: education and job market experience, which I denote as e and n , respectively. The former refers to all education, broadly defined, that would increase a worker's marginal productivity; the latter refers to cumulative labor market experience—i.e., aggregate experience, as opposed to time on the particular job, which has already been captured in t .⁹⁸ We now have:

$$\beta = \beta(e; n); e, n \geq 0. \quad (3)$$

I further assume that the β function has the same properties of monotonic increase and diminishing marginal rate relative to each of its arguments, education and experience, that were reflected in the marginal productivity curve depicted in Figure 1. This means that both education and experience produce positive but diminishing returns for a worker's productivity and, hence—by Equations (1), (2), and the competitive market assumption—on her wage level. Finally, we can make e and n endogenous to the model by having each worker make decisions to affect them based on the other variables within the model, most importantly the wage schedule. This leads to:

$$e = e(w_i); n = n(w_i). \quad (4)$$

This means that an employee's level of education and experience are each functions of the offered wage. Wage levels affect education and experience,

97. See Diesenhouse, *supra* note 89, at 5; see also Henry Hansmann, *When Does Worker Ownership Work? ESOPs, Law Firms, Codetermination, and Economic Democracy*, 99 YALE L.J. 1749, 1783-96 (1990) (discussing governance cost reasons why law firms do not want part-time partners). Pension antidiscrimination laws also may play a role by disfavoring high quality part-time work and creating an incentive to employ either full-time workers or temporary-independent contractors. See generally JOHN H. LANGBEIN & BRUCE A. WOLK, *PENSION AND EMPLOYEE BENEFIT LAW* 162-217 (1990) (discussing pension antidiscrimination laws).

98. Evidence supports the proposition that aggregate experience has some intrinsic relevance to a worker's level of productivity. See BLAU & FERBER, *supra* note 2, at 167-68; GOLDIN, *supra* note 2, at 73, 97; Jacobsen & Levin, *supra* note 49.

each of which in turn affects marginal productivity, which in turn affects the wage, and the cycle continues.⁹⁹

After developing a fuller signaling model, I will add detail to the story. For now, a simpler, more intuitive point emerges. Determining the ultimate direction of the cycle depends crucially on the way that educational and experience profiles respond to the offered wage schedule.¹⁰⁰ Faced with a lower *ex ante* wage schedule, many young married women—indeed, many of the class of women planning on getting married—will have less incentive, compared to men and never married women, to pursue education.¹⁰¹ This lack of education will further lower their salaries, by reducing their intrinsic productivity. Once again, just as was the case with firm behavior in perpetuating statistical discrimination, these decisions will be *rational*. Given the existing wage structure, it will not be cost-effective for some women to pursue education, although they would do so in a first-best world. But the mere rationality of this action does not bless its normative propriety, especially since it is a reaction to an already imperfect market.

99. See, for example, Cain, *supra* note 26, and Gronau, *supra* note 71, for discussions of the difficulties of sorting out cause and effect in analyzing gender discrimination.

100. This point can be seen formally by taking the partial differential of β with respect to the offered wage schedule, w . Consolidating equations (3) and (4) from the text and eliminating the subscripts, we have:

$$\beta = \beta(e(w); n(w)).$$

We are interested in the response of changes in w on β . Thus, we have:

$$\frac{\partial \beta}{\partial w} = \frac{\partial \beta}{\partial e} \frac{\partial e}{\partial w} + \frac{\partial \beta}{\partial n} \frac{\partial n}{\partial w}.$$

But we know, by assumption, that

$$\frac{\partial \beta}{\partial e}, \frac{\partial \beta}{\partial n} > 0,$$

so that the sign of the derivative of β with respect to w will depend on the signs of

$$\frac{\partial e}{\partial w} \text{ and } \frac{\partial n}{\partial w},$$

as well as on the relative magnitudes of all four terms, if these two derivatives have opposite signs, none of which are restricted by the model. These terms behave much like elasticities which can be positive, negative, or zero. A lower wage may *encourage* additional education and persistence, by some combination of income effects and a lower opportunity cost for education, or it may *discourage* each, through a lessened profit opportunity from the increase and a greater substitution effect. Of course, the two may point in opposite directions, and have different magnitudes, etc. We shall consider the general cases where the differential of β with respect to w is negative in this section, and then look at the positive example in the next section.

101. These are those women—presumably quite a few if not the majority—who respond to a lower wage schedule by cutting back on education and taking time off from the job market. The next section considers the case of those who in fact pursue *more* education in the face of these labor market conditions. See Ian Ayres, *Price and Prejudice*, NEW REPUBLIC July 6, 1992, 30, 32 (reviewing EPSTEIN, *supra* note 3) for a quick statement of the view that statistical discrimination will lead to lower investment in education. See also GOLDIN, *supra* note 2, at 214; Coate & Loury, *supra* note 61, at 5 (“[W]hen employers believe workers from a certain group are less likely to be of high ability, they offer them a lower wage. As a consequence, the return from investment in skills is lower to workers from a group so discriminated against. Hence they invest less . . .”); Rose, *supra* note 11, at 443-45 (discussing disinvestment in human capital by discrimination against women).

Further, faced with a lower wage structure both because they are married women and because they are now less educated, and given the other market deficiencies discussed above—particularly taxes and the lack of a viable part-time labor market—these women will become more likely to take stretches off from the job market.¹⁰² Not only will such women face a lower wage schedule in the first instance, because of the employers' guesswork regarding their tenure, but this wage schedule will be further depressed by both the lesser education and the lesser cumulative labor market experience.¹⁰³

All of these factors conspire to create a self-fulfilling prophecy. Once married, women are offered lower wage schedules because of assessments of their likely tenure. This consequence is then compounded by other factors. Taxes push couples to identify a secondary earner, and influence the terms of secondary-earner status. Many women in fact become more likely to take time off than men, a tendency accelerated by the effects of taxation and the deficiencies in the part-time labor market. Married women or young women who want to get married¹⁰⁴ have less incentive to educate themselves for job market skills, and their actual periods of absence from the job market further depress their worth. As in many other areas, life conforms to the predictions of theory. But the test of the initial hypothesis was not random. Small initial discriminations, rationally based, evolved into significantly discriminatory patterns; the test was rigged. Finally, note that the whole story has unfolded thus far without resort to any discriminatory motivation on the part of anyone: there are no women-haters in the picture. We have had only rational firms and individuals, all acting in their own best interests.

D. *Phase III: Slouching Towards Equality*

Many commentators end the story of job market discrimination with some version of the preceding tale. They begin with rational discrimination and illustrate how it becomes a self-fulfilling prophecy. But to end the story at this point leaves these commentators ill-equipped to deal with positive evidence of

102. BLAU & FERBER, *supra* note 2, at 150-54; GOLDIN, *supra* note 2, at 133 ("A married woman was not easily enticed into the labor force by higher wages, but she was, at the same time, encouraged to leave by higher earnings of her husband and other family members."); William Even, *Career Interruptions Following Childbirth*, 2 J. LAB. ECON. 255, 255-56 (1987) ("[Y]oung children have a strong negative effect on the probability of a woman participating in the labor force . . . [A]t least part of the 'wage gap' between the sexes is the result of women having more frequent career interruptions.") (citation omitted); Jacobsen & Levin, *supra* note 49; Maume, *supra* note 17, at 495 ("Weekly child-care payments were a significant predictor of employment turnover.").

103. See, e.g., Jacobsen & Levin, *supra* note 49, at 1-2, 11.

104. Note that the whole pattern has endogenous effects on marriage's desirability, for example, making marriage attractive to many women, because employers pay them less than men in any event, on the assumption that all women will want to marry. See BECKER, *supra* note 68, at 55; Cohen & Haberfeld, *supra* note 10, at 29-31, 42-43 (discussing possible relations between marriage and earnings profiles). On the other hand, for those women committed to the labor force, marriage becomes less attractive because it is an adverse signal.

the gender gap's narrowing. To return to the sets of debates beginning this Article, the interventionists are reduced to arguing with the facts of the optimists, that seem to confirm the non-interventionists' case. If things are really so terrible, why do the objective indicia appear so positive?

We can see that there is not necessarily anything inconsistent between the views of the optimists and those of the interventionists by continuing the story. Nor is there any clear support for the non-interventionist posture in the positive data. Phase II ended with a self-perpetuating, and seemingly viciously downward cycle. All married women faced significantly lower wage schedules, and, accordingly, many cut back on both education and job market persistence. We were headed for an equilibrium in which all married women would be categorically marginalized in the labor force, with ripple effects on all women, whether currently married or not.

The story, however, does not end here. Various factors continue to pull the labor market in a different direction. First, as conditions for married women deteriorate, marginally committed women can be expected to opt out of the labor force. But at some point a reversal is likely to occur. Those women who are left in the work force will be, almost by definition, highly committed ones. Firms will find that women represent a good deal—they may even be paid less than their marginal productivity. But this situation, for standard economic reasons, is not stable in a competitive setting, so wage schedules for women will rise. As they do, many of the effects discussed above will be reversed. This phenomenon might help to explain erratic cycles of discrimination, with significant periods when the trend appears to be favorable for women until a correction in the other direction sets in.

At this stage, signaling also begins to take on special importance. The story told in the prior two Sections had assumed that women formed a homogeneous group. But of course they do not: many women have no intention of having children, or of leaving the work force if they do. The problem such women have had in the story thus far is that they had no way of convincing potential employers of these intentions, and so they were lumped with the rest of women. But as the situation worsens for all women, those intent on having their own careers have increased incentives and opportunities for strategically signaling to firms their greater commitment. For example, by overeducating themselves, such women signal their greater desire to work, and credibly precommit to a strategy of remaining in the work force. Within the model developed above, such education also has a direct, positive effect on their productivity.¹⁰⁵ The market will continue to separate, and for some women this market process may drive them to higher paying jobs with more responsibility.

105. See BLAU & FERBER, *supra* note 2, at 148-49; *Labor Letter: More Schooling Is the Key to Improving Wages of Black Workers*, WALL ST. J., May 11, 1993, at A1.

By extending the story out to this Phase III, we can see how the gender gap might indeed narrow, intermittently or permanently, without changing any of the basic assumptions of the model. But here is the rub: the convergence is not necessarily a happy phenomenon. The gender gap narrows because women only marginally committed to the work force pull out of the race, and those committed to it go to great lengths to indicate their commitment. Quantitative equality is won at a qualitative price. Women act to fit a male pattern of labor, men go on acting as they always have, and a significant opportunity for social change is lost. The promise—or myth—that a greater role for women in the labor market would alter both the work force and the larger society goes largely unfulfilled.¹⁰⁶ We will have slouched toward a predefined, static notion of equality. The next two subsections develop in greater detail the ideas of separating equilibria and signaling behavior that inform this phase of the model.

1. *Separating Equilibria*

I want to introduce a metaphor, drawn from George Akerloff's classic model of lemons in the used car market.¹⁰⁷ I apologize for the unfortunate turn of phrase in this context (and we shall see soon enough that the story of gender discrimination in fact presents a reverse-lemons situation) but the illustration is helpful in advancing discussion. The lemons story in its original form, the used car market, presents an excellent example of equilibrium models in the presence of imperfect and asymmetric information.¹⁰⁸ In that case, the market consisted of two goods: high-quality used cars and lemons. All individuals were assumed to know the proportion of lemons in the market, but only sellers had precise knowledge of which cars were in fact lemons. Buyers made a decision on the basis of the odds of choosing a lemon, and the relative value of the two types of cars. Assume, for example, that two out of five used cars were known to be lemons, that a non-lemon was valued by a potential buyer at \$3000, and a lemon at \$1000. A random car would then have an expected value to a buyer of

$$v = (.4)\$1000 + (.6)\$3000 = \$2200. \quad (5)$$

106. See GOLDIN, *supra* note 2, at 4-9 for a different interpretation of this seeming paradox. Goldin believes that the grip of past discriminatory ideas masks and slows the progress toward equality. I argue, somewhat differently, that these past practices set the terms on which progress is made, casting doubt on the very meaning of progress itself. For other discussions of the disappointment attendant on inflexibility in the face of progress, see SCHWARTZ, *supra* note 7, at 63-91; Czapanskiy, *supra* note 7, at 1451-53; Dowd, *supra* note 7, at 431.

107. George A. Akerloff, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488 (1970).

108. Akerloff himself considered the possibility of no equilibrium, but most of the discussion, and subsequent intellectual history of his article, has focused on equilibrium models, including the limiting case where no market forms. *Id.* at 489-92.

Assuming risk-neutral buyers, either all cars would be sold for \$2200, providing a windfall for the sellers of lemons, or, if the offered price of \$2200 fell below the minimum price acceptable to sellers of non-lemons, all non-lemons would be withdrawn from the market. In this latter case, the price would fall to \$1000, since the probability of a used car being a lemon would go up to one.¹⁰⁹

The pooling of different goods in the first outcome of the lemons model is parallel to the tendency in statistical discrimination to aggregate classes of employees into a common group. But labor markets differ from the used car market in that the very trait of concern to buyers—here the expected tenure of potential employees—is, in several regards, endogenous to the model. In the used car example, whether or not a car was a lemon was exogenous. Only the choice of whether or not to sell was a variable (although I discuss warranties as signals below). This is not true in labor markets. Employees to a significant degree control their likely tenure. The relevant decisions will be affected by the offered wage, which in turn will be related functionally to education and experience and to firms' estimates of predicted tenure.

Consider a simple numeric example. Assume a baseline setting in which married women fall into two classes: those who will work for at least five years, which I take to be the maximum planning horizon of the firm, and those who will only work two. Firms and married women alike know that fifty percent of married women fall into each group. We know from above that the offered wage is a function of expected tenure. Further assuming risk-neutral firms, the employer will predict that a married woman's tenure with the firm will be

$$t = (.5)2 + (.5)5 = 3.5, \quad (6)$$

and will set its wage schedule accordingly. But not all women will react in the same manner to the offered wage schedule. Some will drop out; others will persist. It seems likely that those who persist will have a greater expected tenure than will the entire pool.¹¹⁰

A good amount of empirical evidence appears to support this analysis. Recent studies show that the wage elasticity of working women is lower than that of nonworking ones—in fact, the elasticity of working women is quite close to that for men.¹¹¹ One interpretation of this fact is that a number of factors make the participation decision an especially difficult and all-or-nothing

109. Akerloff assumed inelastic sellers (a fixed supply of used cars). *Id.* Note that it is simple enough to modify the story to have continuous grades of quality.

110. There appears to be a similar separation of African Americans into participating and withdrawn sectors, making wage gap comparisons suspect there, too. See EPSTEIN, *supra* note 3, at 257-59; Bulow & Summers, *supra* note 46, at 399-401.

111. See Triest, *supra* note 74, at 512-13 (summarizing results of econometric study to estimate labor supply elasticities of men and women and comparing results to previous studies).

one. For example, there may be a large after-tax fixed cost of entering the work force, and bearing this cost makes one less elastic, because the costs must be repaid.¹¹² Another interpretation is that working breeds a taste for work—this is an endogenous preference argument of the sort that I have generally sought to avoid in this Article.¹¹³ Yet another interpretation reverses the causal argument: lower wages have driven out the women with higher wage elasticities, and left the pool of working women skewed toward those with lower wage elasticities. That is, working has not changed women, but rather the nature of the workplace—in particular, the wage and tax biases against them—has screened, or sorted, women into two groups.¹¹⁴

All three of these theories are consistent with each other, even mutually reinforcing, and it does not much matter to my thesis which are right or more persuasive. All lead to the same point: women in the work force will be relatively more committed to staying there than will women as a whole, since the latter includes those women not very strongly committed to the work force.¹¹⁵ Other evidence supports the point, too: the actual work experience of women in the work force has risen, among all age cohorts, although the commitment of working women to the work force has always been high.¹¹⁶ Fertility rates have fallen with education and labor force experience, indicating that workers are less likely to have children than are nonworkers.¹¹⁷ In terms of turnover rates and participation patterns across the life cycle, women have moved dramatically closer to historically male patterns.¹¹⁸ These facts show that the labor force commitment of working women has been increasing with work, whereas child-rearing decisions and responsibilities have declined.

The process of driving out less committed women means that the percentage of committed women in the work force changes, in a direction opposite the typical lemons effect. Instead of making inferior products more likely, the dynamic here makes less committed women less likely to remain in the labor force. As this becomes clear, firms update their probable tenure

112. See Jerry A. Hausman, *Labor Supply*, in *HOW TAXES AFFECT ECONOMIC ACTIVITY* (Henry V. Aaron & Joseph A. Pechman eds., 1981) (data suggesting that wives have low wage elasticity when fixed cost of work is taken into account). For a discussion of the effects of the fixed cost of work as it relates to the popularity of lotteries, see Edward J. McCaffery, *Why People Play Lotteries, and Why It Matters*, 1994 WIS. L. REV. (forthcoming); Aloysius Siow, *Fixed Costs of Work and the Demand for Lotteries*, 29 ECON. LETTERS 299 (1989).

113. See JON ELSTER, *SOUR GRAPES: STUDIES IN THE SUBVERSION OF RATIONALITY* (1983); Vicki Schultz, *Telling Stories About Women and Work: Judicial Interpretations of Sex Segregation in the Workplace in Title VII Cases Raising the Lack of Interest Argument*, 103 HARV. L. REV. 1750 (1990); Cass R. Sunstein, *Legal Interference with Private Preferences*, 53 U. CHI. L. REV. 1129 (1986).

114. See Bulow & Summers, *supra* note 46; Salop & Salop, *supra* note 57, at 620.

115. In this regard, it is relevant that 80% of nonworking women are married, while only 65% of working women are married. Marriage itself has served as a screen. SORESENSEN, *supra* note 2, at 74.

116. BLAU & FERBER, *supra* note 2, at 162-64; GOLDIN, *supra* note 2, at 41.

117. BLAU & FERBER, *supra* note 2, at 262-64; GOLDIN, *supra* note 2, at 139-42; LINDEN, *supra* note 2, at 4-5.

118. BLAU & FERBER, *supra* note 2, at 74-80; STATISTICAL ABSTRACT, *supra* note 15, at 387 (Table No. 618); Goldin, *supra* note 17, at 24-25.

assessments, and salaries for women will start to rise. This process has been borne out by evidence showing that while women are still more likely to quit than men, the turnover rate of women in the work force has been dropping dramatically over the last twenty years, and is approaching that of men.¹¹⁹ Women who do work, work the same way men do. Married women's salaries will rise with both the promise of greater stability (via the employer's prediction) and the fact of greater persistence (via the endogenous effect of work experience on productivity). Single women will benefit as well, since the threat of marriage becomes less worrisome for an employer. Again, these effects reinforce one another. Of course, as salaries rise, incentives will change: more women will be drawn back into the work force from among the ranks of the less committed, and those in the work force will have less need to signal their own commitment.

Similar self-selection effects can be expected to flow from the type of gender-specific wage-tenure profiles noted above in Figure 4. Firms may react to the statistically lesser persistence of women by setting a steeper wage schedule, that starts out below the male equivalent but rises more rapidly. In the face of such schedules, women would self-select: those unlikely to remain for a long time would not enter, seeing only the below-market initial pay. In contrast, not only would likely long-term employees be attracted by the ultimate wage prospects but also, once they were on the job, their incentive to stay would actually be *greater* than that for equivalently employed men, since their rate of salary increase would be steeper. In these instances, it is an explicit institutional structure, rather than the general price level, that brings about the separation effect and leads employed women to act as men traditionally have.

While economic theory generally predicts instantaneous calibration of all these effects so as to produce an immediate stable equilibrium, the real world is never so precise. Firms lack full information as to the preferences of potential employees, the signaling activity takes time to unfold, and even the employees themselves may be unable to predict their future with accuracy.¹²⁰ The phenomena described above are likely to play themselves out in cycles: periods during which women catch up, followed by periods of backlash caused by too many less committed women entering the market.¹²¹ In the long run, these very oscillations may be expected to result in an upward trend, as lesser committed workers periodically take time off, and are forever thereafter stamped with the mark of their own lack of persistence. In fact, this is what appears to be happening in the real world: the narrowing of the gender gap

119. BLAU & FERBER, *supra* note 2, at 74-80.

120. See, e.g., GOLDIN, *supra* note 2, at 154-57, 216-17; Schultz, *supra* note 113.

121. Posner makes a similar point. Posner, *Economic Analysis*, *supra* note 3, at 1323.

represents a trend extending back before the present century, but one marked by many periods of consolidation and even regression.¹²²

2. *Signaling Behavior*

Up to this point in the story, I have assumed that the married women who are my primary subject have done little to differentiate themselves on the basis of their degree of labor market commitment, except by actually persisting on the job. These women have reacted to market forces, by deciding whether or not to work or to have children, and so on. Now I want to enrich the model by considering the possibility of married women's engaging in signaling behavior.

Michael Spence produced the classic work on market signaling in the mid-1970's.¹²³ The motivation for the model was an informational asymmetry in the labor market similar to the one we have been exploring: employers are concerned with the innate abilities of potential employees, but the employees have very little incentive to tell the truth about these. Spence looked at the role of education in solving the dilemma. He assumed that education added nothing to an employee's intrinsic productivity, but instead merely signaled the employee's innate ability, which he took as immutable. The key assumption needed for the model to work was that the costs of education are negatively correlated with ability, i.e., it costs less for more able workers (including psychic costs) to acquire a given level of education than it costs less able ones. Given this assumption, Spence was able to show that, under a wide range of situations, a signaling equilibrium emerges.

To continue this background for a bit, consider again the lemons problem. In the classic formulation of this model, no signaling equilibrium exists: the market does not separate out by quality. Instead, if we return to the numeric example given above, either the market forms a common pool at the \$2200 level or the market breaks down, and only lemons are offered. But what if the owners of non-lemons had some way to convince potential buyers of the worthiness of their cars? Perhaps such owners could credibly warrant the quality of their goods. Such agreements would not give rise to any *actual* liability—we are, by assumption, dealing with non-lemons—but the signals are needed to convey the information. Sellers of lemons would not offer warranties because, for them, a warranty would carry real costs. The signal should be effective. The market will attain a signaling equilibrium, where sellers of non-lemons could be paid the value of their car, less the costs of signaling.

Signaling has much relevance for gender discrimination. In the model I have presented, productivity is not an immutable characteristic (or a function

122. See GOLDIN, *supra* note 2, at 213; SORESENSEN, *supra* note 2, at 15, 17.

123. SPENCE, *supra* note 67.

of an immutable characteristic, such as ability), but varies with education and work experience.¹²⁴ The key variable in the model is expected tenure. Women can signal, or at least attempt to signal, their degree of commitment to the work force. One way to signal is by not getting married or not having children, and women are indeed marrying later and having fewer children as the gender gap closes.¹²⁵ But these are awkward signals, at best. Employers may not believe that a young woman is not married (in part because they are aware of the phenomenon of married women hiding their wedding rings on the way to job interviews¹²⁶); they may feel constrained, or be legally restrained, from asking if a woman is married; they may assume that any unmarried woman is soon to be married; they may even think that unmarried women are unstable or disruptive.¹²⁷ Similar problems arise with not having children. There is no very good way to convince a potential employer that one will not have children at some time. Job market persistence itself signals future job market persistence, and there is some evidence that working women might stay at their particular jobs longer than men.¹²⁸ But this strategy only works once one has a job, and tends to have the unfortunate effect of locking one into that

124. As an aside, the use of the terms "mutable" and "immutable" characteristics in the signaling literature is not free of normative import. Spence himself did not use the word "signal" when dealing with immutable characteristics such as race or gender, preferring instead the term "index." But the realities are more complex. Immutable characteristics such as race or gender are not important in themselves, at least in my model; they might be for models featuring tastes for discrimination, or, as in some of Richard Epstein's examples, rational discrimination based on intrinsic differences between races or genders, or economies of scale in segregationist policies. In the rational discrimination model presented here, the immutable characteristic of gender is important as a predictor for a mutable trait, either ability or, more specifically to this model, commitment. A woman cannot change her sex, at least without incurring considerable costs, but she can change her degree of commitment to the labor market. Indeed, getting married itself is an act that sends out the wrong signal on this score—that is, it does for women—and thus the evidence that married women hide their wedding rings prior to job interviews is not surprising. SCHWARTZ, *supra* note 7, at 9-26. The evidence supports the fact that marriage increases men's wages while having a lesser effect on women's. See Bulow & Summers, *supra* note 46, at 400 (citing to Francine Blau, *Discrimination against Women: Theory and Evidence*, in *LABOR ECONOMICS: MODERN VIEWS* (William A. Darity ed., 1984)).

125. See LINDEN, *supra* note 2, at 1, 4.

126. See SCHWARTZ, *supra* note 7, at 9-26.

127. Note that married men do not have these problems because, under contemporary patterns, being a married male sends out a positive signal of stability and commitment, whereas being an unmarried male may send out an adverse signal. Note also that the textual analysis in part explains my focus on married women—firms view all women as either actually or potentially married. In any event, almost 90% of all women have married at least once by age 40. LINDEN, *supra* note 2, at 4.

128. A factor that might influence this persistence is the wife's lesser mobility. See Norman Bonney & John Love, *Gender and Migration: Geographical Mobility and the Wife's Sacrifice*, 39 SOC. REV. 335 (1991). Also note that there is an important difference, from the firm's point of view, between an employee's leaving for another job, which men tend to do, and having to stay home, as women tend to do. In the former case, the employee is being lured away by a more competitive job offer, which the employer is theoretically free to match. Women do have higher labor market turnover rates than do men. BLAU & FERBER, *supra* note 2, at 79.

job. The best signal for married women remains Spence's own example: education.¹²⁹

Evidence shows that the narrowing of the gender gap has been associated with an increase in the education of women, although the gap has narrowed within all educational classes.¹³⁰ In the present model, education has three distinct effects. First, it is a pure signal—not only of ability, as in Spence's model, but also of commitment. An advanced professional degree, for example, shows the rational employer that the young woman is thinking ahead many years, and has made an investment in education that could only have been *ex ante* rational if she intended to capitalize on it. Second, education is not merely a signal, but also a means of credible pre-commitment to the persistence strategy. This is especially true where the education has been financed with debt, as is typically the case, at least for professional degrees: the costs will be sunk, behind the employee, and will need to be paid. This ongoing liquidity need can be a way of convincing employers that the woman is likely to stay on the job; it converts the signal from one of *ex ante* thinking to one of *ex post* need.¹³¹ This credible pre-commitment strategy is akin to the idea of firms that have market power building up excess capacity as a defense against new entrants.¹³² In both cases, participants incur an otherwise irrational cost solely to convince other market participants of the probability of future action. Finally, in our model, education also has an intrinsic effect on productivity.

For all of these reasons, women committed to the labor market are likely to educate themselves, perhaps even excessively or inefficiently, and this process of education has self-fulfilling elements. The need to signal is a measure of inefficiency; signaling is only needed once the first-best conditions of perfect and complete information do not exist. Signaling through education is also inefficient because both the job market and education display significant indivisibilities; one cannot pick and choose precise levels of either. Instead, an individual who wants to signal through education must typically complete a degree requirement in order to obtain a full-time position. As with statistical discrimination, all of the decisions about signaling will be rational—it will only make sense to signal if the benefits of signaling exceed its costs—but this is not the same thing as saying that signaling is efficient.¹³³

Tying this into the point on the separating equilibria, the inducement toward education is likely to vary directly with the extent of the wage bias against women. When a large bias exists, the opportunity costs of pursuing

129. I am not suggesting either that education has value only as a signal or that it only signals commitment. Education has direct benefits and signals ability as well as commitment, as in Spence's analysis. Thus, men will continue to educate themselves too.

130. BLAU & FERBER, *supra* note 2, at 191-95; SORESENSEN, *supra* note 2, at 94-96.

131. See quote from GOLDIN, *supra* note 2, cited *supra* note 48.

132. See, e.g., B. Curtis Eaton & Richard G. Lipsey, *Capital, Commitment, and Entry Equilibrium*, 12 BELL J. ECON. 593 (1981) (modelling phenomena and discussing theoretical literature).

133. SPENCE, *supra* note 67, at 152-76.

education are low at the same time as the benefits of signaling for the committed group can be great. We should expect relatively more education to take place, *ceteris paribus*,¹³⁴ among women committed to the labor market in times of great discrimination. The market among women is separating because the lower wage schedule has divergent effects on women's incentives to educate themselves—committed women have greater incentives, lesser committed women have reduced ones. As an aside, this means that individual examples often perceived as "success stories"—so-and-so made it in a man's world, in large part by working herself through school—may be interpreted more as evidence of discrimination than as optimistic stories of the possibility of success through education or of the absence of barriers based on gender.

In terms of the facts, the gender gap's narrowing has indeed been accompanied by a steady, and disproportionate (to men) rise in the education of women.¹³⁵ A good deal of the apparently favorable changes in the work force have come from professional jobs, characterized by advanced degrees. For example, women have been making very significant inroads in computer programming, law, medicine (as physicians), accountancy, and engineering.¹³⁶ And, as the analysis above suggested, women have become relatively *less* likely to educate themselves as the gender gap has narrowed, as it has been since 1970 or so.¹³⁷ Signaling behavior has had real effects, and has worked together with a seemingly intrinsic ebb and flow in discriminatory patterns.

E. *Summary of Model*

The model began by assuming rational firms and individuals, a competitive market, and equal marginal productivity between men and women, holding education and labor market experience constant. From these humble beginnings, transaction costs, imperfect information, taxes and incomplete markets began systematically to distort the situation, first through the seemingly benign device of rational discrimination. Over time, these factors had profound effects, and led to a situation in which married women in particular faced a sizeable discrimination in the work force. There were ripple effects on other women, as well, as firms viewed all women as at least potentially married. Market outcomes could no longer lay claim to being efficient, in a strong, first-best sense.

134. I assume no discrimination in education, and a market open to rewarding educated women, i.e., there are no entry barriers such as marriage bars. See GOLDIN, *supra* note 2, at 160-79.

135. GOLDIN, *supra* note 2, at 215-16; LINDEN, *supra* note 2, at 4-5.

136. LINDEN, *supra* note 2, at 5; SORENSEN, *supra* note 2, at 62.

137. SORENSEN, *supra* note 2, at 2, 101-02. African Americans' wage and educational gains seem to be following a similar course. Large gains in the first years of affirmative action have given way to small gains or even no gains at all in recent years. See James P. Smith, *Affirmative Action and the Racial Wage Gap*, 83 AM. ECON. REV. (PAPERS & PROC.) 79, 81-84 (1993).

In order to capture something of the dynamism of real-world economic effects, the model proceeded through three phases. In the first phase, real-world wage schedules and search costs led to a need for firms to estimate in advance the likely tenure of potential employees. Incomplete information then led these firms to conclude that women, especially young married women, were likely to have relatively short tenures. In the second phase, this prophecy began to fulfill itself. Offered lower wages, women in fact began to withdraw from the work force, a movement hastened by a tax system slanted against two-earner households and a market lacking in high quality part- and flexible-time opportunities. Human capital decisions followed, and played a continuing role as both cause and effect. In the third phase, the pattern reversed itself, as low wages drove out the lesser committed women. At the same time, women more highly committed to the labor market began to signal their commitment through later marriage, fewer children, and more education. The possible overeducation resulting from these incentives had not only signaling effects, but also intrinsic effects on productivity and the need for money income.

The moral of this story is that market failures, especially in a dynamic setting, are central to market outcomes. Small initial discriminations, even if entirely rational, can easily lead to a situation wherein severe discrimination is rampant. This can all happen as a result of intrinsic, endogenous features of markets working themselves out in the presence of real-world imperfections, without any extra-market assumptions or dark conclusions about human nature. It is important to see that, in the particular case of gender discrimination in the work force, the effects of market failure over time cannot be seen by looking at the gender gap alone. The gender gap has in fact narrowed, but on terms strikingly similar to the predominantly male labor patterns of centuries. The market has not changed; people—primarily women, who have borne the brunt of the induced behavioral changes—have changed to accommodate it. Women are finally being treated like men—the gender gap is closing—because they have decided, under some considerable pressure, to contort their behavior to fit into a man's world. This outcome, it would seem, violates the normative intuitions of neoclassical theorists regarding the political morality of the market. The balance of this Article turns to these normative aspects of the story.

IV. THE TROUBLE WITH DOING NOTHING

The model in Part III was meant to demonstrate that market failures are at work in the story of gender and labor markets, and that these market failures lead the real world away from the ideal one of first-best theory. Standard economic analysis holds that market failures justify an inquiry into the possibility of market intervention. There is, however, a deep skepticism among most practitioners of law and economics—to be sure, a skepticism grounded

in experience—over the wisdom or efficacy of interventions in a second-best world.¹³⁸ This skepticism forms the basis for a weaker, and more plausible, claim from the anti-interventionists: it is not that the rationality or efficiency of firm actions is an automatic bar to intervention, or that intervention is *never* helpful, but rather that markets are apt to work their way to a better answer, even in the presence of market failures, than government intervention can provide. It is a second-best answer: given market imperfection, governmental intrusion is likely to cause more harm than good. I intend to challenge this view, in the particular case of gender in the work force, in this Part.

Richard Epstein is perhaps the most noted exponent of the anti-interventionist position, although he is by no means the only one. In *Forbidden Grounds*, Epstein advocates dismantling virtually the entire range of employment discrimination laws, and he does so, at least partially, in the name of efficiency.¹³⁹ If discrimination is a problem (or, rather, if the effects of past discrimination on today's workers are a problem),¹⁴⁰ one might be tempted to address it through distributive tax and welfare policies.¹⁴¹ This approach is ostensibly consistent with the liberal spirit of the two classic welfare theorems: the first theorem holds that free markets reach Pareto optima, and the second holds that any particular Pareto optimum can be reached through a suitable redistribution of initial endowments.¹⁴² In this view, the market is the ideal allocative mechanism—the perfect vehicle to aggregate individual preferences—and government regulation can only muck things up. The only acceptable vehicle for change is pure, direct redistribution. This view, however, ignores or downplays the decisive role of market failures, especially as iterated over time.

138. See, e.g., COASE, *supra* note 42, at 26-31, 181-85; EPSTEIN, *supra* note 3, at 12, 34, 87, 266; Fischel & Lazear, *supra* note 3, at 906-13; Coate & Loury, *supra* note 61, at 1-4, 31-33; Posner, *Efficiency and Efficacy*, *supra* note 3, at 516-21. For a useful summary and discussion of the deregulatory movement, see Clifford Winston, *Economic Deregulation: Days of Reckoning for Microeconomists*, 31 J. ECON. LITERATURE 1263 (1993).

139. EPSTEIN, *supra* note 3, at 23-24, 26, 59-60, 76-78. Epstein argues that the government should not regulate private markets; in the case of employment discrimination laws, he in fact advocates doing something—eliminating the laws. But I take this to be semantically consistent with a “do-nothing” label. For reviews of Epstein's book which focus on its utilitarian or efficiency-oriented themes, see Ayres, *supra* note 101; John J. Donohue III, *Advocacy Versus Analysis in Assessing Employment Discrimination Law*, 44 STAN. L. REV. 1583 (1992); Samuel Issacharoff, *Contractual Liberties in Discriminatory Markets*, 70 TEX. L. REV. 1219 (1992); George Rutherglen, *Abolition in a Different Voice*, 78 VA. L. REV. 1463 (1992); J. Hoult Verkerke, *Free To Search*, 105 HARV. L. REV. 2080 (1992).

140. Many commentators, in addition to Epstein himself, have argued that irrational or inefficient discrimination cannot persist in competitive markets, so that the only “problems” are due to past discrimination. See, e.g., BECKER, *supra* note 28, at 16-17; Fischel & Lazear, *supra* note 3; Posner, *Economic Analysis*, *supra* note 3.

141. I initially read Epstein as taking exactly this position. See EPSTEIN, *supra* note 3, at 12. In his letter to me, *supra* note 3, Professor Epstein insisted that he meant no such thing. I therefore attribute the position—certainly a coherent one, played out in the various political proposals for vouchers for education, food stamps, and the like—to a generic observer.

142. See BOADWAY & BRUCE, *supra* note 84, at 3; KREPS, *supra* note 84.

This Part takes particular aim at the anti-interventionists' do-nothing case. To come to terms with this case, it will help to understand why, when, and to what extent efficiency might be a value. This may seem like an unwarranted divergence from the primary analysis, but I ask the reader to bear with me, largely because efficiency norms are so often asserted or abandoned without any explicit treatment of what is meant by efficiency or why we should care about it at all. I therefore explore briefly what "efficiency" can mean. In the following Part, I will use this discussion to advance and justify a policy reform proposal, featuring partial deregulation and a significant change in the tax laws.

The first Section of this Part contains a discussion of the various positive meanings of efficiency. The second Section navigates through the diverse normative claims involved in efficiency analysis in an imperfect world. I return more directly to the gender discrimination story in the third Section, sketching the perils of the do-nothing approach. The fourth Section argues that redistribution alone is not enough to cure the troubles in the work force. The fifth Section adds a few thoughts on analyzing claims for "equality" in the context of gender discrimination in the work force.

A. *What Is Efficiency?*

When commentators talk of efficiency, they generally invoke one or more of three basic concepts: Kaldor-Hicks, Pareto superiority, or Pareto optimality.¹⁴³ All three are concerned with the allocation of resources, as opposed to their distribution.

The Kaldor-Hicks standard holds that a given allocation of resources is efficient if the gains to the winners are large enough such that the winners could compensate the losers and still be better off. It is a sort of social cost-benefit analysis, and perhaps the most common use of efficiency in the law and economics literature.¹⁴⁴ When society decides to build a prison in Community X, against the better interests of X's own citizens but for the greater good of the entire society, it has acted in a Kaldor-Hicks efficient manner. Kaldor-Hicks does not require any actual compensation to be paid; if such compensation were paid, the standard would become Pareto superiority, discussed below. Because the Kaldor-Hicks standard, like the Paretian ones, grew out of attempts to avoid interpersonal utility comparisons, the norm is

143. See Jules L. Coleman, *Efficiency, Utility, and Wealth Maximization*, 8 HOFSTRA L. REV. 509, 550 (1980); Gary Lawson, *Efficiency and Individualism*, 42 DUKE L.J. 53, 84-92 (1992); Richard Posner, *The Ethical and Political Basis of the Efficiency Norm in Common Law Adjudication*, 8 HOFSTRA L. REV. 487 (1980).

144. See, e.g., RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 13 (3d ed. 1986) ("When an economist says that free trade or competition or the control of pollution or some other policy or state of the world is efficient, nine times out of ten he means Kaldor-Hicks efficient, as shall this book."). I disagree with Posner's strong statement.

strictly and properly an ordinal one.¹⁴⁵ This means that Kaldor-Hicks is *not* utilitarian, at least in any of the most common variants of utilitarianism; it is instead *wealth* maximizing.

A simple example illustrates these points.¹⁴⁶ Assume that individual utility is given by the square root of wealth—this is a common illustrative utility function, because it captures diminishing marginal utility quite nicely. Further assume that the social welfare function is simply additive: we determine the aggregate social utility by summing the unweighted utility of each individual. There are two persons, Rich and Poor. Now consider two states of the world. In State I, Rich has 81 units of wealth, and Poor has 25. Social utility is 14 (9 plus 5, the square roots of each person's wealth). In State II, Rich has 100 units of wealth, and Poor has 9. State II is Kaldor-Hicks superior to State I, because the winner *could* compensate the loser *out of her gains* and make both parties at least as well off. That is, since Rich has gained 19 units (100 minus 81), she could compensate for Poor's loss of 16 units (25 minus 9) and make both parties at least as well off as they were in State I. But State II is *not* a utility improvement over State I; utility has fallen from 14 to 13 (10 plus 3). These numbers are summarized in Table 1.

State	Rich's Wealth	Poor's Wealth	Total Wealth	Utility
I	81	25	106	14
II	100	9	109	13

TABLE 1. *Utility Versus Wealth Maximization*

The reason for the difference is that Kaldor-Hicks ignores wealth effects, or the varying utility of wealth, altogether. It collapses into wealth-maximization—State II has more total wealth, 109, than does State I, 106.¹⁴⁷

The Paretian efficiency conditions arose out of the work of the turn-of-the-century economist, Vilfredo Pareto, and his pre-Kaldor-Hicks attempt to avoid utilitarian ethics' reliance on interpersonal cardinal utility comparisons.¹⁴⁸ Pareto superiority exists when a given action improves the welfare of at least one person without harming any other person. Pareto superiority does not

145. Lawson, *supra* note 143, at 89-90.

146. I am very grateful to my colleague Tom Griffith for suggesting this example and for straightening me out on the distinctions involved.

147. Note that this statement is true whether or not money is involved. Kaldor-Hicks efficiency insists on maximizing wealth, however it happens to be defined or measured. Money income is simply the most common approach to the measurement problem.

148. Posner, *supra* note 143; G. Busino, *Vilfredo Pareto*, in 3 THE NEW PALGRAVE: A DICTIONARY OF ECONOMICS 799 (John Eatwell et al. eds., 1987).

necessarily involve any interpersonal comparisons; it simply requires individuals to make ordinal rankings between states.¹⁴⁹ Absent externalities, including envy—and any cognitive errors that might lead individuals to act contrary to their own self-interest—free trades or exchanges are Pareto superior; no one is made worse off by the trade, and one person is presumably made better off. If I buy a car for \$10,000, for example, neither I nor the car's seller is worse off, and at least one of us is typically better off. Free markets enable Pareto superior trades. Were we to require fair compensation under a Kaldor-Hicks standard, we would arrive at Pareto superiority, since the compensated party (by definition of fair or full compensation) would be no worse off for the action. Thus, Kaldor-Hicks efficiency is sometimes called potential Pareto superiority.¹⁵⁰

Finally, Pareto optimality is a description of end states. It consists of all those states where no one individual could be made better off without making at least one individual worse off. If we were to turn all individuals loose in some figurative marketplace, and give them complete knowledge of what all of the other agents possessed, a Pareto optimum would theoretically be reached; each agent would continue trading until she could find no one willing to make another trade on terms at least minimally acceptable to each. The marketplace idea is, however, only one of many means to reach a Pareto optimum—i.e., through a series of Pareto superior trades. A Pareto optimum need not be reached through Paretian moves.¹⁵¹ To take a common, if overdrawn, example from existing literature, it is possible to reach a Pareto optimum by confiscating all of society's wealth and giving it to a single greedy individual; no further trades would be possible that did not take at least something away from this single Midas.

Under perfect market conditions, all three efficiency-related concepts converge. This has led to a tremendous amount of confusion in both the positive and normative uses of efficiency in real-world policy analysis, where they often diverge. Under ideal conditions, parties reach a Pareto optimum through a series of Pareto superior moves. Each move is also Kaldor-Hicks efficient, since in fact the potential Pareto superiority has been actualized. The winners from trade *had* to compensate those who would otherwise be losers, because no one would consensually lose in a trade. Indeed, this also follows from neoclassical economic assumptions and definitions, here of rationality and

149. *But cf.* DENNIS C. MUELLER, PUBLIC CHOICE II 400-07 (1989); Y.K. Ng, *The Possibility of a Paretian Liberal: Impossibility Theorems and Cardinal Utility*, 79 J. POL. ECON. 1397 (1971).

150. *See, e.g.*, Posner, *supra* note 143, at 491 (noting that Kaldor-Hicks efficiency is "sometimes called 'Potential Pareto Superiority.'").

151. *See, e.g.*, Coleman, *supra* note 143, at 548 ("One can secure Pareto-optimal outcomes through Pareto-superior, non-Pareto-superior, or Kaldor-Hicks-efficient steps. In other words, we can reach a Pareto-optimal result in a particular case by making no individuals worse off or by making some individuals worse off.").

revealed preference.¹⁵² Resources flow naturally to their highest and best use. Every individual optimizes subject to endowment constraints. Free trade and rational self-interest lead to an *optimum optimorum*. There is a triple efficiency, as exclusively Kaldor-Hicks efficient and Pareto superior trades have led to a Pareto optimum end state.

There is much that is seductive in this idealized notion of competitive, general equilibrium, and a good deal of efficiency analysis is concerned with matching the structural conditions of this perfect state to the real world. Examples are the principles that marginal benefits equal marginal costs, that a factor's price equals its marginal rate of transformation, and that prices equilibrate each consumer's marginal rate of substitution for the given good over all others.¹⁵³ It is also interesting to note that, under perfect markets, there is no need for efficiency analysis at all, since what happens is definitionally efficient, a basic lesson from the Coase theorem.¹⁵⁴ It is only under the imperfect conditions of the real world that efficiency analysis can have prescriptive, as opposed to merely explanatory, force. In practice, however, it becomes critical to specify both the standard of efficiency and its underlying norms, because these can and do diverge in the real world in a way that they do not in theory. Without such a deep and complex inquiry, it is too easy to fall into the categorical position that efficiency dictates doing nothing at all.¹⁵⁵ This is, indeed, the position of Epstein and others when it comes to gender discrimination, but I mean to suggest that it is a wrong one, even on efficiency grounds.

B. *Why Do We Care About Efficiency?*

Discussions of the normative propriety of efficiency too often seem to begin with some definition of efficiency in mind. If, for example, we begin with Paretian standards, we are unlikely to get very far in addressing problems of gender discrimination, because Paretianism insists that no one be hurt by a change, a condition rarely achieved in real-world social transformations. I want to begin further upstream, by asking in the first instance for a general political or moral philosophy within which efficiency might play a role.¹⁵⁶ My inquiry begins with the question, Why should we care about efficiency at all? This methodology will yield a benchmark for sorting through the ambiguities

152. See, e.g., KREPS, *supra* note 84, at 42-43, 140-43.

153. See, e.g., VARIAN, *supra* note 92, at 11-12, 82-93, 219, 225-27.

154. See COASE, *supra* note 42, at 10-16 (updating his views on the problem of social cost); Robert C. Ellickson, *The Case for Coase and Against "Coaseanism,"* 99 YALE L.J. 611 (1989); George J. Stigler, *Two Notes on the Coase Theorem*, 99 YALE L.J. 631 (1989).

155. For criticism of efficiency analysis as traditionally employed, see Guido Calabresi, *The Pointlessness of Pareto: Carrying Coase Further*, 100 YALE L.J. 1211 (1991); Lawson, *supra* note 143.

156. I am here taking what has been called a "top-down" approach. Cf. JULES L. COLEMAN, *RISKS AND WRONGS* 6-13 (1992) (discussing top-down versus middle-level theory).

involved in any discussion of the meanings and uses of "efficiency." It turns out, not surprisingly, that our reactions to inefficiencies depend critically on why we care about efficiency in the first place.

I posit three basic sets of reasons why one might care about efficiency: utilitarianism, liberalism, and libertarianism. I have chosen these three theories in particular because they capture much of the real, apparent, or possible support for some efficiency norm. I admit immediately that I cannot advance these discussions very far within the confines of this Article, and my conceptions of each political philosophy are, by necessity, rather thin. Still, I explore these ideas because some analysis is essential to my broader project of showing how efficiency analysis cuts in the case of gender discrimination in the work force.

1. *Efficiency and Utilitarianism*

I begin with utilitarianism because it is often thought, or presumed, that the underlying bases for efficiency norms are utilitarian. Indeed, many commentators seem to confuse efficiency with utilitarianism.¹⁵⁷ But, as the technical definition of Kaldor-Hicks indicates, this identification is not accurate. While there are many versions of utilitarianism—involving cardinal or ordinal, total or average utility, etc.—most utilitarians accept some variant of declining marginal utility of money income.¹⁵⁸ As long as they do so, Kaldor-Hicks efficiency does not correspond with utilitarianism; the association can only be made in the implausible case that all individuals have the same utility of wealth, and that this utility is constant. This is a particularly severe problem if, as Posner claims, nine times out of ten policy analysts mean Kaldor-Hicks when referring to efficiency norms.¹⁵⁹ There are, however, serious reasons to doubt the accuracy of Posner's claim.¹⁶⁰

157. See, e.g., Jennifer H. Arlen, *Reconsidering Efficient Tort Rules for Personal Injury: The Case of Single Activity Accidents*, 32 WM. & MARY L. REV. 41, 42 (1990) ("The Pareto criterion differs, therefore, from the efficiency that the 'law and economics' literature customarily employs, *total social utility maximization*, which depends on the ability to aggregate utility.") (emphasis added) (citations omitted); Richard A. Posner, *Utilitarianism, Economics, and Legal Theory*, 8 J. LEGAL STUD. 103 (1979) (discussing, and attempting to sever, the link between economic analysis and utilitarianism, in favor of wealth maximization as an ethical norm; this becomes a strong theme for Posner over the next several years). But cf. RONALD DWORKIN, *A MATTER OF PRINCIPLE* 275-89 (1985) (criticizing Posner's occasional conflation of efficiency and utility); Coleman, *supra* note 143, at 512 ("[W]hile some relationships between efficiency criteria and utilitarianism exist, these relationships are not necessarily justificatory in nature.").

158. See EPSTEIN, *supra* note 3, at 260; POSNER, *supra* note 144, at 11; McCaffery, *supra* note 112.

159. POSNER, *supra* note 144, at 13.

160. Most economists and political-economists (as opposed to legal academicians) proceed using Paretian notions of efficiency. See, e.g., BOADWAY & BRUCE, *supra* note 84, at 2-3; VARIAN, *supra* note 92, at 225-26; Joseph Stiglitz, *Pareto Efficient and Optimal Taxation and the New Welfare Economics*, in 2 HANDBOOK OF PUBLIC ECONOMICS 991, 992 (Alan J. Auerbach & Martin Feldstein eds., 1987) ("Economists, in their role as economists, should limit themselves to identifying Pareto efficient allocations . . ."); see also DWORKIN, *supra* note 157, at 237-40 (distinguishing between wealth maximization and efficiency); JOHN RAWLS, *A THEORY OF JUSTICE* 66-70 (1971) (defining efficiency as Pareto Optimality); SEN, *supra* note 6, at 136 (arguing that Pareto efficiency is almost certainly the most widely used criterion

In any event, Kaldor-Hicks efficiency might have at least a limited role to play under a utilitarian philosophy. It may create a type of presumption in favor of utility-enhancement, providing a rule of thumb. Unless we have decisive reasons to believe that there are significant wealth effects involved, that is, that the winners have a lower marginal utility of wealth than the losers, the Kaldor-Hicks superior transaction should go forward.¹⁶¹ For example, there would appear to be little reason not to correct costlessly (if possible) an externality between a presumably risk-neutral firm and its risk-averse neighbors, under the Kaldor-Hicks principle. Risk-neutral firms ought to follow privately a Kaldor-Hicks, cost-minimizing path. Perhaps most importantly, in the wide range of situations where we cannot say, a priori, which side of a given transaction is likely to have a greater marginal utility of money, there are utilitarian reasons to go with a Kaldor-Hicks answer. This might explain the real, utilitarian power of the ex ante Pareto superiority move, common in fields such as contract and tort analysis, where risks are imposed reciprocally and the incidence of harm is uncertain.¹⁶² In other words, efficiency might serve as a type of rule, as opposed to act, utilitarianism, specifying how general classes of transactions should be structured.

There also seems to have been a more or less informal move to restate the Kaldor-Hicks test in utile terms. A given transaction is Kaldor-Hicks efficient if the gainers could compensate the losers in *utiles* and all parties would be at least as well off as before the transaction took place. This redefinition reverses the answer given above in Table 1. The basic move seems necessary, in part, to avoid the so-called Scitovsky paradox, wherein two states could be Kaldor-Hicks efficient relative to each other.¹⁶³ But it is a move that, by definition, completely collapses Kaldor-Hicks efficiency into utilitarianism, and thus misses the point of the Kaldor-Hicks test, which is to avoid interpersonal utility comparisons.¹⁶⁴ Nonetheless, many uses of efficiency in the law and economics literature in fact seem to come down to the idea that the benefits of a putative action outweigh its costs, taking into account a wide range of nonmonetary costs and benefits. Generally, such discussions ought to proceed by substituting "utility-improving" for "efficient," although one senses that,

in modern welfare economics and is typically taken to be entirely noncontroversial).

161. See generally Robert D. Willig, *Consumer's Surplus Without Apology*, 66 AM. ECON. REV. 589 (1976) (demonstrating when observed elasticities can be decisive for utilitarian analysis). But see George W. McKenzie, *Consumer's Surplus Without Apology: Comment*, 69 AM. ECON. REV. 465 (1979) (criticizing Willig).

162. See, e.g., Arlen, *supra* note 157; George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537 (1972); Posner, *supra* note 143.

163. See MUELLER, *supra* note 149; Coleman, *supra* note 143; Ng, *supra* note 149 (discussing Sen's impossibility theorem that Paretianism is inconsistent with liberalism).

164. See Lawson, *supra* note 143, at 89-90; see also Posner, *supra* note 143, at 491 (offering a hypothetical using *utiles*). See generally *supra* note 147 (discussing money income as measure). Kaldor-Hicks efficiency analysis can go forward using any objective measure of value.

given prevalent intellectual fashions, this more accurate statement would somehow be less appealing.

On the other hand, Paretian ethics have considerable appeal for almost any type of utilitarian; it is difficult to object to the mutually beneficial trades of the Pareto superiority standard.¹⁶⁵ Paretian ethics, however, do not get us very far in the real world. Paretian optimality is a description of end states that has little independent appeal; it is a necessary, but far from sufficient, condition for a social optimum. It is not sufficient, because of the manifest unattractiveness of certain Pareto optimal states. Further, it is only necessary because there is no reason to leave potential Pareto superior moves on the table; it is Pareto superiority that provides the heart of any operative Paretian concept of efficiency.

Pareto superiority does have appeal to the utilitarian. But it is important to see that a commitment to Pareto superiority, as part of a political philosophy, can have one of two senses. In a *positive* sense, it means a commitment to seeking out and facilitating all Pareto superior moves. This version seems unobjectionable, especially to utilitarians. Indeed, utilitarians are likely to believe in the efficacy and normative propriety of free markets simply because they offer the most practical way of increasing utility, given the difficulties of revealing preferences. In the quite different, *negative* sense, however, a commitment to Pareto superiority means opposing any *non*-Paretian moves. This is the sense that libertarians such as Epstein tend to invoke, as I discuss below, in the case of gender discrimination. This idea is antithetical to utilitarianism. In its strongest form, utilitarianism is opposed to all rights and entitlements. It is difficult to make a utilitarian case for giving a veto to Paretian concerns, which would elevate virtually every aspect of the status quo to a right.¹⁶⁶

Efficiency has a limited role to play under a pure utilitarian philosophy. Kaldor-Hicks efficiency is not utility-enhancing, except under very special circumstances that most utilitarians reject. Kaldor-Hicks may be a rough approximation, or a useful rule of thumb, but it adds no normative content to utilitarianism per se. When many authors speak of efficiency, they in fact seem to mean utility-enhancing; in this, somewhat unfortunate sense, utility and efficiency are connected. But here, too, the real concern is exclusively with utility. Paretian norms have appeal, but only in the positive sense of a commitment to seeking out mutually beneficial trades or activities; they cannot, consistent with utilitarianism, exert a veto over all social change. The result is

165. But see SEN, *supra* note 6, at 136.

166. For an illustrative discussion of the connections between rights and utilitarianism, see T.M. Scanlon, *Rights, Goals, and Fairness*, in CONSEQUENTIALISM AND ITS CRITICS 74 (Samuel Scheffler ed., 1988). For an attempt to mediate rights and utilitarianism, largely through the device of rule-utilitarianism, see David Lyons, *Utility and Rights*, in THEORIES OF RIGHTS 110 (Jeremy Waldron ed., 1984).

that efficiency is a concept that is neither necessary nor even particularly helpful to utilitarians.

2. *Efficiency and Liberalism*

If efficiency's normative appeal rests on its attenuated and largely mistaken connection to utilitarianism, it seems odd that it should be so popular. After all, utilitarianism has been in decline for many years, and most contemporary philosophers disparage it.¹⁶⁷ Does the success of efficiency analysis mean that utilitarianism retains a strong populist appeal, notwithstanding its virtual banishment by the literati? I believe that another answer is possible: efficiency analysis is not rooted in utilitarianism at all, as the prior discussion in fact implied, but rather derives its appeal from characteristically liberal concerns such as personal autonomy and neutrality.¹⁶⁸

Contemporary economists sometimes seem to forget that market theory had its historical genesis in liberal political economics. Adam Smith, David Ricardo, and John Stuart Mill were all, to varying degrees, classical liberal supporters of market mechanisms.¹⁶⁹ These early political economists were led to efficiency norms out of a solicitude for individuals and their preferences, a fact that has been obscured by diverse uses and meanings of the term "efficiency." But once we invoke efficiency in the real world, we must choose among competing underlying philosophies and positive meanings, and a tension between instrumental and individualistic justifications for efficiency begins to emerge.

The efficient world of pure market theory is one that respects individuals. They are allowed and even encouraged to engage in trade, and all benefit. One sees this clearly in the work of Adam Smith. One of his central ideas was that purely selfish actions of uncoordinated individuals can increase the opportunities for all.¹⁷⁰ Virtually all liberal thinkers give *some* room to

167. See, e.g., Coleman, *supra* note 143, at 510-11 ("Utilitarianism, however, has been the target of powerful, and, in the minds of many philosophers, decisive objections."); Posner, *supra* note 157, at 106 ("Today legal theorists who discuss utilitarianism tend to reject it as a basis of normative legal theory.").

168. For representative contemporary discussions of liberalism, see, e.g., DWORKIN, *supra* note 157, at 181-204; WILL KYMLICKA, *LIBERALISM, COMMUNITY AND CULTURE* (1989); CHARLES E. LARMORE, *PATTERNS OF MORAL COMPLEXITY* (1987); STEPHEN MACEDO, *LIBERAL VIRTUES: CITIZENSHIP, VIRTUE, AND COMMUNITY IN LIBERAL CONSTITUTIONALISM* (1990). For a powerful contemporary view of liberalism, see JOHN RAWLS, *POLITICAL LIBERALISM* (1993).

169. In a passing reference, Rawls calls Smith and David Hume utilitarians. RAWLS, *supra* note 168, at xiv. In some regards, this label is accurate—specifically insofar as it refers to a psychological view of individual motivation and action. In other regards, however, it is misleading—if, for example, it is taken to mean that aggregative claims trump individual ones in all instances. Part of my aim in this Article, which will ultimately require more extensive treatment elsewhere, is to explore the non-utilitarian bases of market philosophies.

170. ADAM SMITH, *AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS* (R.H. Campbell et al. eds., 1976) (1776); Jerry Evensky, *Ethics and the Invisible Hand*, 7 J. ECON. PERSP. 197 (1993) (discussing, *inter alia*, ADAM SMITH, *THE THEORY OF MORAL SENTIMENTS* (D.D. Raphael & A.L.

markets, where Paretian efficiency can hold sway. Rawls, for example, defines efficiency as Pareto optimality, and leaves ample room for markets to operate.¹⁷¹ Dworkin countenances Paretian trades as part of a social commitment to deep equality.¹⁷² Efficiency, in this liberal sense, is not just about consent. It is also fundamentally concerned with opportunity: a commitment to efficiency involves giving more people more choices about more things, and respecting those choices in the face of potential paternalistic intervention. As Dworkin comments, there is no reason not to let Paretian trades go forward. Liberal efficiency is an individualistic, not a social, norm.¹⁷³

In addition to serving liberals' concern for autonomy, efficiency and markets in general satisfy a liberal desire for neutrality.¹⁷⁴ To understand this claim, it is critical to interpret neutrality as entailing a *process*, a particular style of dialogue. Relying on efficiency norms seems to allow us to avoid making contestable, or transparently non-neutral judgments. In a wide range of cases, we can support efficiency-enhancing moves because, *ex ante*, it appears that no particular group or set of interests will be hurt, and some will be helped. So long as the diverse effects of a law appear to us to be random or incidental—not particularly affecting any specified group of persons or interfering with the pursuit of any particular vision of the good life—the law seems to be as neutral as we can hope for it to be. But this can only be the first step in a neutrality dialogue.

The next step in the process of seeking neutrality involves recognizing that supposedly neutral principles may in fact have decidedly non-neutral effects. Consider, for example, general principles of taxation. We might begin with the idea that everyone should pay the same dollar amount of tax—a lump sum, poll, or head tax. We justify this tax on neutral principles, believing, like Hobbes, that everyone benefits equally from government.¹⁷⁵ Soon enough,

Macfie eds., 1976) (1759)).

171. RAWLS, *supra* note 160, at 66.

172. DWORKIN, *supra* note 157, at 269-70.

173. Cf. BOADWAY & BRUCE, *supra* note 84, at 2 (importance of individualism to welfare economics). Note that Kaldor-Hicks efficiency, with its emphasis on aggregate wealth, tends to lead to more of a social norm.

174. See, e.g., Bruce Ackerman, *What Is Neutral About Neutrality?*, 93 ETHICS 372 (1983) (defending concept of neutrality against critics). Now, neutrality is (and I believe rightly) a much battered term these days. But, properly employed, neutral concepts play a valuable role in a liberal dialogue; for example, a notion of neutrality seems very much in play in Rawls' requirement for public reason to inform political action. See RAWLS, *supra* note 168, at 212-54. In particular, an insistence that rules must be justified without explicit reference to any particular group interests or any specific visions of the good life can help shape liberal political discussions in much the same manner as the more familiar burden of proof concept mediates political dialogue. See, e.g., LARMORE, *supra* note 168, at 40-68. I have benefitted from conversations with Larry Solum on this and related points. See Lawrence B. Solum, *Constructing an Ideal of Public Reason* 11 (May 28, 1993) (unpublished manuscript, draft on file with author).

175. See THOMAS HOBBS, *LEVIATHAN* 226 (Michael Oakeshott ed., 1946) (1651). See generally Richard Musgrave, *A Brief History of Fiscal Doctrine*, in *HANDBOOK OF PUBLIC ECONOMICS*, *supra* note 74, at 1, 16-18 (discussing benefit doctrine of taxation).

however, we come to see this principle as not being neutral at all; either we question the factual predicate or the normative propriety of equal benefit taxation. We then move to "ability to pay" norms, and develop a system of progressive marginal rates based on some quasi-utilitarian theory of equal sacrifice.¹⁷⁶ This progressive income tax now strikes us as more neutral than the poll tax. We are well aware, of course, that a progressive tax system affects different people differently, although we try to minimize the differing effects through formal concepts such as horizontal equity, with its mandate to treat like cases alike.¹⁷⁷ But neutrality is not really intended as a formal or substantive criterion; instead, it serves a rhetorical, justificatory role. As noted above, as long as the diverse effects of law appear to us to be random or incidental, the neutral outcome seems as good as we can do. The head tax had violated this rhetorical test, because it seemed to fall too harshly on the poor—to ignore wealth effects altogether, that is—but the progressive rate formula seems acceptable.

Now, what I want to do is to elevate the dialogue one more level, to show how the supposedly neutral principles of the actual tax system are indeed having distinct behavioral and distributive effects, on specified groups (women) and distinct versions of the good life (dual-earner marriages). This is the type of claim that liberals must consider, and, if it is persuasive, must look for alternative practices, just as with the move from head taxes to more deeply individuated ones. A similar progression, for example, has led to the debates over affirmative action for minorities, and about how to evaluate test data in educational admissions policies, and so on.¹⁷⁸

There is, of course, a perfectly fair complaint that neutrality in any form is doomed to be chimerical; we could abandon it altogether. But, like Churchill's famous observation about democracy, there is not necessarily any better alternative.¹⁷⁹ It is critical once again to understand that neutrality in a justificatory sense entails a *process*; there can be no slavish, static devotion to semantic norms. Liberal neutrality involves a commitment to a particular style of dialogue. It is in precisely this justificatory sense of neutrality that some have invoked concepts of efficiency or the market. This, too, strikes me as only an opening position, subject to rebuttal. As in the tax example, we must be open to the charge that any particular type of efficiency analysis, or

176. See Musgrave, *supra* note 175, at 18-21 (discussing ability to pay doctrine); WALTER J. BLUM & HARRY KALVEN, JR., *THE UNEASY CASE FOR PROGRESSIVE TAXATION* 64-68 (1953); HENRY SIMONS, *PERSONAL INCOME TAXATION* (1938).

177. For critical discussion of the horizontal equity norm, see Louis Kaplow, *Horizontal Equity: Measures in Search of a Principle*, 42 NAT'L TAX J. 139 (1989). But cf. Richard A. Musgrave, *Horizontal Equity, Once More*, 43 NAT'L TAX J. 113 (1990).

178. See Kelman, *supra* note 31.

179. "Indeed, it has been said that democracy is the worst form of Government except all those other forms that have been tried from time to time." Winston Churchill, House of Commons, Nov. 11, 1947, in OXFORD DICTIONARY OF MODERN QUOTATIONS 55 (1991).

efficiency analysis in any particular context, is not neutral. This point is especially important in the real, imperfect world, where efficiency analysis, although it might share a common, deep structure, is not necessarily determinate: we can make different efficiency arguments for opposing outcomes in a range of cases. Part of what I aimed to achieve in the model I presented in Part III was to show inefficiencies at work where others have seen only or mostly efficiency.

In any event, I believe that a liberal commitment to efficiency rests not on the wealth-maximizing property of Kaldor-Hicks efficiency, but on such values as fairness, neutrality, decentralized authority, and autonomy-enhancement. When inefficiencies appear, as measured by deviations from the idealized structure of pure, competitive general equilibrium conditions, the liberal ought to scrutinize these to see if the presumptive fairness of markets is being violated. If it appears that particular groups or visions of the good life are being impacted, and if a change that—all things considered—seems to make matters better, in the sense of being fairer, or more like the impartial outcomes of first-best theory, is evident, then the liberal ought to make the change. An extreme case is where formal barriers simply close markets to certain individuals or groups; it is easy to see why such markets should be opened. A variant of this liberal perspective is subjective efficiency—the view, common among the Austrian school of economists, that efficiency only has meaning relative to a specified group or goal.¹⁸⁰ Using this tool, once we are led—here, by the awareness of market failure itself—to assist a given group such as women, we next ask how most efficiently to accomplish the goal. But this subjective efficiency, while related, fails also to ground the reason for the intervention on efficiency terms. Under a liberal theory of efficiency, as I mean it, both the reason and the means of intervention are informed, if not dictated, by efficiency analysis. In the case of gender discrimination, for example, inefficiencies are a sign that something is indeed wrong, and efficiency analysis can suggest possible changes. I shall pursue this liberal sense of efficiency further in Part V.

3. *Efficiency and Libertarianism*

Finally, I consider libertarianism, an outgrowth of classical liberalism and the favorite philosophy of Epstein. When libertarians use efficiency, they tend to invoke Pareto superiority in its exclusively negative sense.¹⁸¹ They veto

180. See Israel M. Kirzner, *Philosophical and Ethical Implications of Austrian Economics*, in *THE FOUNDATIONS OF MODERN AUSTRIAN ECONOMICS* 75-88 (Edwin G. Dolan ed., 1976). For recent discussions of Austrian economics in the legal literature, see Lawson, *supra* note 143, 96-97; Linda B. Schwartzstein, *Austrian Economics and the Current Debate Between Critical Legal Studies and Law and Economics*, 20 *HOFSTRA L. REV.* 1105 (1992).

181. See, e.g., EPSTEIN, *supra* note 3, at 24-25.

moves that hurt anyone; uncompensated takings are strictly prohibited. It is easy enough to see how this idea grows out of the libertarian's commitment to individual autonomy and a minimalist state, but this view of efficiency has all of the problems of a strictly negative view of liberty, to which it is related. Such a view ignores altogether the problems of inequitable distribution and iteration. Worse, in the present context, the negative view of efficiency precludes any move designed to remedy market failures.

Consider, for example, the presence of an externality—a firm emitting smoke that damages its neighbors' lungs and crops. The entitlement rule is settled and favors the neighbors. Under perfect market conditions, the neighbors either collect damages from the firm, which goes on polluting, or, if it is Kaldor-Hicks efficient—wealth maximizing—the firm cuts back on its polluting. This is an illustration of the Coase theorem. Now suppose that there are market failures—the neighbors lack information, or the transaction costs of getting together are prohibitive. The firm continues, quite rationally, to emit smoke. This is efficient in a Kaldor-Hicks sense, given the real-world costs,¹⁸² but we are far from the conditions of competitive equilibrium. In particular, there is an ongoing Pareto inferiority at work: each time the firm pollutes, it hurts the neighbors.

Why should the libertarian accept this market failure? If she takes the status quo as a baseline, the libertarian must concede that regulating the firm is not a Paretian move, because the firm will suffer. On the other hand, if the libertarian takes the perfect market outcome as the baseline, as she generally does,¹⁸³ she would have to concede that the status quo involves an ongoing, Pareto inferior move, and she would have to permit regulation in this instance. But then the door would be open to *any* government move that seeks to correct for market failure and better approximate the consensual outcomes of perfect markets. Market failures undercut the morality of the market: it is no longer perfectly consensual, because there are ongoing winners and losers. The battle must be waged on second-best grounds—where, to be sure, the libertarian has powerful, but no longer theoretically decisive, arguments.

The story demonstrated by the model in Part III was that market forces and failures led to restricted opportunity sets for women. A combination of transaction costs, informational problems and taxes resulted in women being paid less than men and having to alter their behavior to fit the patterns of a male-dominated work force. To be sure, there is no classic externality involved, but market failures are still operating to harm some individuals vis-à-

182. Guido Calabresi has observed that a Pareto optimum *with transaction costs* "is what is," because, by definition, any pure Paretian improvements would be made voluntarily. Calabresi, *supra* note 155, at 1216 ("[I]f Pareto optimality means a place where no improvement can be made without ex ante creating the possibility that there will be some losers, then we *are always* there.").

183. See, e.g., EPSTEIN, *supra* note 3, at 105.

vis others.¹⁸⁴ So long as these market failures persist, the libertarian ought to consider the possibilities of a cure. Whether any particular action will make things better or worse is no doubt a difficult question; the important thing to realize is that it is a question we must answer. Libertarianism cannot deliver a preemptive first strike against the inquiry.

C. *The Case Against Doing Nothing*

The preceding discussion has established a basis for criticizing the non-interventionist approach to gender discrimination: efficiency is not a logically dependent proxy for doing nothing. This and the next subsection consider in greater detail Epstein's argument that we should abolish all employment discrimination laws and, if there is a need for some action on corrective justice grounds (it is unclear that Epstein thinks there is), that we should address it through distributive mechanisms.¹⁸⁵ Epstein's approach turns out not to be Kaldor-Hicks efficient, utility-enhancing, or consistent with a richer liberal sense of efficiency. It is at best predicated on the weak, negative version of the Pareto superiority standard that flows from and into libertarianism.

The evolution of market forces presented in Part III has placed us in a deeply second-best setting.¹⁸⁶ If we were only to eliminate the employment discrimination laws, we would not be moving very far in the right direction. Transaction costs would still abound. The government would still be very much in the picture, via tax policies at least. Markets would not be complete. Perhaps worst of all, the dynamic effects of market imperfections would still be in place; human capital decisions would have been made, and the probabilistic base for statistical discrimination would not have changed. The most that could be said, on efficiency grounds, is that, from the point of eliminating employment discrimination laws onward, Pareto superiority would obtain.¹⁸⁷ But Pareto superiority (or some close approximation) is a very weak justification for a social order. In contrast, a significant change in the taxing regime can increase both utility and wealth while comporting better with the characteristically liberal intuitions behind efficiency norms.

184. There is, of course, legitimate room to dispute the contours and even existence of market failure. Posner, *Economic Analysis*, *supra* note 3, at 1320-21, has a more limited sense of market failures at work in the story of gender discrimination than I do, limiting them to monopsonistic exploitation of higher relocation costs, aggression against women, and ignorance of the average qualities of women due to information externalities. As noted above, I would add taxes and transaction costs in contract design at least, as well as search costs, incomplete (as opposed to imperfect) information, and incomplete markets.

185. EPSTEIN, *supra* note 3, at 12, 257-66.

186. The classic citation for this principle, which holds that, in a world where all of the conditions for an ideal are not met, one cannot necessarily improve matters by simply eliminating distortions, is R.G. Lipsey & Kelvin Lancaster, *The General Theory of Second Best*, 24 REV. ECON. STUD. 11 (1956-57).

187. EPSTEIN, *supra* note 3, at 259-65 (discussing distributive winners and losers from Title VII). The move to repeal the laws is not *itself* Pareto superior, however, as Epstein himself concedes.

To return to the specific case of gender discrimination, the dangers of doing nothing in the current setting are clear. Rational discrimination is both a cause and an effect of the problems. It is an effect because informational deficiencies created it, but it is also a cause because it has led to an altered incentive structure that has hurt women in particular and helped to make the workplace impervious to change. Firms and other institutional actors may be acting rationally, or efficiently, but this is not enough. Efficiency must ultimately be justified by its effects on individuals: efficiency is good if the wealth it creates improves the lot of individuals,¹⁸⁸ or if efficient markets respond to individual preferences. Epstein ironically commits a version of the sin for which he so often condemns others: he forgets to respect individual autonomy. By focusing only on firms, on the institutional or demand side of the market, where he finds no monopoly power or irrational behavior, Epstein fails to see that individuals, on the supply side, are not being allowed to optimize, to pursue their vision of the good life as best as their endowments might allow. Where individuals are blocked by market failures from pursuing their life plans, efficiency talk loses much of its normative appeal.

Imagine the consequences of repealing the employment discrimination laws, without doing more. Wages ought to fall for women. If they did not, then the antidiscrimination laws would not have been constraining, and hence not inefficient. If the employment discrimination laws were generating productive inefficiencies, as Epstein maintains, then total productivity would increase, causing the total wage level to rise. This might then be a Kaldor-Hicks efficient move, but not necessarily. If the repeal facilitated more statistical discrimination, and if such discrimination worked against a more elastic class, then wealth might fall under the change; the efficiency gain would be private, to the firms, not public.¹⁸⁹ The lower wages would drive some women out of the market altogether, and would lead others to signal their greater worth; we would have a fresh round of the dynamics sketched in model in Part III, now played out in the face of greater discrepancies in wages.

At the same time, all that I have described in Part III would still have transpired. Repealing Title VII cannot undo the history and effects of market failures. Married women have already separated into two groups, one having taken great steps to prove its commitment to work, the other not. How would eliminating Title VII affect this separation? Women who have withdrawn from the work force (or who never entered it)¹⁹⁰ would still face barriers in convincing potential employers of their commitment should they decide to

188. See, e.g., Posner, *supra* note 143, at 497 (noting that wealth-maximization is "consistent . . . with a desire, rooted in the principles of autonomy and consent, to minimize coercion."); Richard Posner, *Wealth Maximization Revisited*, 2 NOTRE DAME J.L. ETHICS & PUB. POL'Y 85 (1985).

189. Bulow & Summers, *supra* note 46, at 398-401; Schwab, *supra* note 40, at 232-33.

190. Note that the analysis of Part III extends to those women who have never worked in the first instance, due to market failures.

reenter (or enter), and to make matters worse, such women have already sacrificed on education and labor market experience. Taxes would still form a wedge between marginal productivity and take-home pay, and the wedge would be greater for married women who find themselves—as most would continue to do—the secondary, or marginal, employee in two-earner households.¹⁹¹ There is no compelling reason to believe that part-time or flexible-time positions would become more plentiful. In short, there is no very good reason to believe that we would be closer to achieving a normatively attractive efficiency or equality. Such a society might be Pareto superior in the narrow sense that firms would no longer be forced to do things against their own perceived self-interest, but this hardly makes the resultant state efficient in any normatively compelling sense. Indeed, if this is all that is meant by efficiency, then eliminating all forms of coercion, public or private, would get us there—efficiency would collapse into libertarianism, just as some have collapsed it into utilitarianism—and Epstein would have a tautological efficiency defense to add to his moral or natural law arguments for liberty.

Clearly, more is going on when economists and lawyers talk about efficiency than is implicit in the libertarian agenda. Efficiency is not simply a code word for laissez-faire policies. One can view the market forces examined in Part III, for example, as inefficient in several senses. Kaldor-Hicks and utilitarian improvements are almost certainly possible. Further, we are far removed from the deep structure of a competitive general equilibrium: marginal costs do not equal marginal benefits, and so on. Any serious economist would admit to inefficiency in such a context, whether or not she would think that an intervention were available to make things better. The fact that any possible corrective would likely be non-Pareto superior is not dispositive of the argument from efficiency.

Of course, the deeper question, common to all problems of the second-best, is how we should structure matters given that perfect market conditions, and their triply efficient outcomes, are not readily forthcoming. Epstein chides others for comparing flawed markets to perfect government interventions, but his criticism cuts both ways—he openly compares flawed interventions to perfect markets.¹⁹² I leave this set of issues to Part V, after I have explained why distributive policies would not suffice to eliminate, and might even aggravate, gender bias in the work force. The basic thought in this subsection is that doing nothing, after years, even centuries, in which rational individuals have responded to market failures, is not an especially compelling policy, even on efficiency grounds alone. It cheapens whatever normative force inheres in

191. See LINDEN, *supra* note 2, at 2, 9.

192. EPSTEIN, *supra* note 3, at 78 ("It is a serious mistake to compare imperfect markets with perfect legislative systems.").

efficiency norms to say that efficiency has been achieved wherever individuals and institutions act rationally, without coercion.

D. *Why Redistribution Is Not Enough*

Even strong libertarians such as Robert Nozick and Richard Epstein occasionally concede that efficiency alone is not enough.¹⁹³ Epstein, for example, turns to the distributive policies countenanced by the second welfare theorem, which holds that if one wants to reach a different point on the Paretian frontier, she need only redistribute initial endowments. Epstein suggests that, to the extent we are concerned with gender discrimination, we should address it through standard redistributive policies: "a combination of tax and welfare systems, keyed to individual wealth."¹⁹⁴

There are, however, severe problems with the redistributive solution in this context. When market failures are not at work, redistributive policies may be more efficient than regulatory approaches, because redistribution can be targeted to the appropriate beneficiaries without significantly skewing market incentives.¹⁹⁵ But where incentives are already skewed—as Part III argued is the case in the market for women's labor—redistributive policies simply do not suffice to address the problem. Redistribution tries to compensate the victims of market failure without curing the market failure itself. Such redistribution may even aggravate the market failure, further restricting individuals' choices. Gender discrimination, for example, is a deep and complex issue, not a simple function of wealth.

Suppose we were to follow Epstein's suggestion and repeal the antidiscrimination laws, replacing them with a more generous welfare system designed to benefit historically discriminated-against groups (including women), and financed out of increased income taxes. How would this impact the decisions facing women, especially married women, in the work force? To the extent such policies are keyed to individual wealth, they would give a

193. See, e.g., ROBERT NOZICK, *THE EXAMINED LIFE* 30-33 (1989) (discussing need to limit power to make bequests).

194. EPSTEIN, *supra* note 3, at 12.

195. A standard example—certainly familiar to Epstein—is rent control, which inefficiently distorts the market pricing signal, and therefore can be expected to lead to lesser quantities and qualities of rental housing, and to black markets among tenants. Society would be better off giving housing subsidies or lump sum grants to the needy and letting the market work itself out on the basis of the new distribution. See Richard A. Epstein, *Rent Control and the Theory of Efficient Regulation*, 54 *BROOK. L. REV.* 741 (1988); see also Robert Moffitt, *Incentive Effects of the U.S. Welfare System: A Review*, 30 *J. ECON. LITERATURE* 1 (1992). Note that such an example implies a partial equilibrium view, which looks at isolated markets. For example, it does not consider distortions in the markets affected by the requisite taxes. A general equilibrium perspective, which would take into account the effects on all markets, would also be considerably more complex. See generally RICHARD A. MUSGRAVE & PEGGY B. MUSGRAVE, *PUBLIC FINANCE IN THEORY AND PRACTICE* 249-62 (5th ed. 1989) (discussing partial and general equilibrium models).

bonus for not working and hence place an added penalty on working.¹⁹⁶ Employers, anticipating all of this, would rationally conclude that women were likely to persist less than men, and the cycles mentioned in Part III would increase in amplitude. Women married to high earning men would probably not be affected at all,¹⁹⁷ except perversely, by a higher tax that would increase the disincentive for them to work. But such women form a key part of the labor market and the typical employer's probabilistic base.¹⁹⁸ The market for women's labor would remain incomplete; the welfare policy could even continue to discourage part-time work.

Even if we were to modify this proposal by effecting the redistribution through a lump sum payment—which virtually all economists would view as more efficient than redistribution keyed to individual wealth¹⁹⁹—it is hard to see how the dynamic problems chronicled in Part III would change for the better. Indeed, if women are motivated to work by income or wealth effects, a lump sum grant might reduce their commitment to the work force, increasing their general labor supply elasticity. This would alter employers' probabilistic information, and once again the market dynamics I discussed in Part III would play themselves out. In other words, redistribution alone could make things worse.

The second welfare theorem holds that if the right conditions—complete and free markets, rational actors, perfect information, and so on—are in place, then redistribution can alter the particular Pareto optimum that is reached. This important principle explains why a distributive theory of justice might occupy a role in a complete liberal political theory relying on markets: given perfect market conditions, society is not forced to choose between distributive concerns and Pareto optimality. Without perfect market conditions, however, we cannot achieve distributive justice through redistribution alone. We are giving money where what is needed is enhanced opportunity. An efficiency-driven dialogue naturally leads us to this conclusion. In Part V, I consider quite a different set of policies—different from both the current regulatory regime and from mere redistribution—for altering the structure of the labor market. First I want to say a few words about equality.

E. *Equality of What?*

The preceding discussion has focused on how to define efficiency and how to make the labor market more efficient, particularly with respect to its impact

196. See Moffitt, *supra* note 195.

197. This is because the transfer to them as women would be offset by the higher tax rate applied against their husbands.

198. See McCaffery, *supra* note 78, at 1028-29; see also LINDEN, *supra* note 2, at 2, 10-11 (noting high prevalence of working wives at upper income levels).

199. See, e.g., MUSGRAVE & MUSGRAVE, *supra* note 195, at 279.

on women. Prevalent intellectual fashion continues to speak of an equity-efficiency trade-off, as though the two norms existed in constant tension.²⁰⁰ But there are in fact very important, structural ways in which efficiency concerns relate to equitable concerns.²⁰¹ The critical question for equality analysis is the choice of space, in Sen's phrase, for measuring equality.²⁰² Popular discussion about gender discrimination, as well as a good deal of scholarly effort, has focused on the wage gap between men and women, or other objective indicia such as relative participation rates in various occupations. The normative problem is that such measures do not offer an especially rich sense of primary goods, as understood by Rawls and others.²⁰³ Nor do they assist us in considering capabilities, or functionings, where Sen and others would move the debate.²⁰⁴ Money is not everything.

The model developed in Part III has left us with a world in which women are given two basic choices: act like men have always acted in the workplace, or stay home. Given this limited choice, and the myriad ways in which the institutions of society have made it constraining, it is not surprising that the gender gap should be closing. Once working women begin to prove that they have accepted the ultimatum to act like men, they are paid like men. But such a structure hardly affords an especially rich sense of equality to women or men, in terms of opportunities, or respect, or freedoms to pursue varying life plans—what really matters to most of us. It is small wonder, then, that many feminist critics have turned against equality analysis altogether,²⁰⁵ and that the statistics of objective improvement continue to belie feelings of subjective unease. Efficiency is very much relevant to these paradoxes, because efficiency is precisely about the ability to convert certain resources into others that might better further one's life plans. The market is efficient only when people can trade freely; when they can convert their resources into any other resources of equal social value.

When women are unable to find attractive work-family combinations, or to work at all except under general patterns set up a century ago during a time of entrenched patriarchy, the corresponding inefficiencies lead directly to inequalities relative to women's capabilities. Imagine a marketplace with a very

200. See, e.g., ARTHUR M. OKUN, *EQUALITY AND EFFICIENCY: THE BIG TRADEOFF* (1975). See DWORKIN, *supra* note 157, for criticism of trade-off language.

201. See SEN, *supra* note 6, at 7-8, 10, 136-41.

202. *Id.* at 12-16.

203. See, e.g., RAWLS, *supra* note 168, at 178-90; RAWLS, *supra* note 160, at 90-95; G.A. Cohen, *Equality of What? On Welfare, Goods and Capabilities*, in *THE QUALITY OF LIFE* 9 (Martha C. Nussbaum & Amartya Sen eds., 1992); John Rawls, *Social Unity and Primary Goods*, in *UTILITARIANISM AND BEYOND* 159 (Amartya Sen & Bernard Williams eds., 1982). Ronald Dworkin focuses on resources, in an expansive sense, as the measure of equality. See Ronald Dworkin, *What Is Equality? Part 1: Equality of Welfare*, 10 PHIL. & PUB. AFF. 185 (1981); Ronald Dworkin, *What Is Equality? Part 2: Equality of Resources*, 10 PHIL. & PUB. AFF. 283 (1981).

204. SEN, *supra* note 6, at 4-5, 39-55.

205. See, e.g., Mary E. Becker, *Prince Charming: Abstract Equality*, 1987 SUP. CT. REV. 201 (1988); Patricia A. Cain, *Feminism and the Limits of Equality*, 24 GA. L. REV. 803 (1990).

limited number of goods: the former Communist regimes of Eastern Europe provide handy examples. Everyone is equal in her ability to buy these goods, and everyone would presumably be offered an equal price. We can even press the thought experiment further, by providing everyone with equal initial endowments. The market limitations still have very different impacts on people's life plans. Those interested in the offered goods are helped; those uninterested in them, but yearning for some other type of good, are out of luck. The current job market is precisely such a limited market. It offers only one type of job, differentially impacting those with different preferences and life plans. The choices of family structure are similarly limited.

That women may one day be paid like men for acting like men represents a very hollow, static notion of equality. Just as we ought to carefully consider the various meanings and rationales for efficiency, we ought to look into the meaning and rationales for equality. The danger of focusing on such objective indicia as the gender gap is that we might get exactly what we have asked for—only to be left even more unhappy and confused in the end.

V. GROPING FOR SOLUTIONS

The model developed in Part III illustrates the problem that market failures, playing themselves out over time, have led women to change their behavior while facing diminished opportunity sets. Women have been given two basic choices: prove that they can act like men and be treated as such, or drop out. As a result, the gender gap among working women has narrowed even while the structure of the work force remains little changed. A stunning shift in the demographics of the labor market has been accompanied by only the most minor changes in its institutional fabric.²⁰⁶ The gender gap has narrowed, but little else has changed.

In light of this set of problems, Part IV criticized the do-nothing approach advocated by Epstein and others on efficiency grounds. No normatively powerful concept of efficiency should compel society to accept the distortions wrought by decades, even centuries, of discrimination. At the same time, efficiency analysis suggested that mere distributive policies are not enough to address the structural problems brought on by the economic cycles spinning out of market failures. The distributive answer gives a static solution to a dynamic problem; it leaves the opportunity sets of women as restricted as ever, and possibly more so. Indeed, the very failure of the distributive answer confirms that market failures are at work.

206. The shift in demographics is chronicled in BLAU & FERBER, *supra* note 2, at 29-32, 258-92; GOLDIN, *supra* note 2, at 10-57; and Gunderson, *supra* note 15, at 46. The lack of changes in the institutional structures of the market is discussed in GOLDIN, *supra* note 2, at 179-83; Dowd, *supra* note 7, at 431.

So the question becomes, What is to be done? At this point in the story many would abandon efficiency norms altogether, as at best indeterminate and at worst inconsistent with liberal principles. I argue instead that efficiency norms are sufficient to justify intervention and, further, that such norms can help to fashion particular remedies. My proposed solution has two parts: the first prong proposes substantial deregulation, the second, a significant change in the tax law. Together, as I explain more fully below, these proposals should increase both utility and wealth.

The first part represents a partial agreement with Epstein: we should drop the regulatory prohibition against unequal pay, although I believe that we should retain the categorical aspects of Title VII that deal with discrimination in hiring and promotion. The idea behind this prong of the proposal is simple: if markets work better with women getting paid less, at least in the present, so be it.²⁰⁷ There is much to be said against standing in the way of the rational decisionmaking processes of firms, whether in the labor market, the housing market, or elsewhere.²⁰⁸ For example, if indeed women pose higher risks to firms because of their lesser persistence, and should therefore rationally be paid less—and if they are not less productive on average, Title VII's equal pay requirement is not itself constraining—then two related effects can follow from the equal pay requirement. One, we are penalizing exactly those firms that “do the right thing,” by hiring women and paying them equally despite women's lower ex ante value. Such firms will probably still have to pay men as much as other firms do, and will be paying an implicit tax in the form of the greater payments to women. Two, and related, we are rewarding firms that can escape Title VII's implicit redistribution or tax, by successfully isolating women into certain job categories, moving to less heterogeneous workplace areas, switching over to capital-intensive production, and so on. In important regards, the means chosen by Title VII to address discrimination might be counterproductive vis-à-vis its own apparent ends.

207. For arguments why Title VII might be efficient, see Donohue, *Title VII*, *supra* note 4. One reason that Title VII might be efficient is that it solves some of the collective action problems raised by the phenomenon discussed in Part III. For example, if women and other minorities are convinced that firms will not be able to practice even rational discrimination against them, they could alter their human capital decisions accordingly. See GOLDIN, *supra* note 2, at 156 (discussing expectations of women). But cf. Spade & Reese, *supra* note 1 (discussing contemporary expectations of women and men college students). Firms might also develop strategies for developing employee capital, knowing that the antidiscrimination laws will require similar investments by others. See BLAU & FERBER, *supra* note 2, at 130-32 (discussing example of the Corning Inc. office).

208. See, e.g., Robert D. Cooter, *Market Affirmative Action*, SAN DIEGO L. REV. (forthcoming 1994). I saw a draft of Professor Cooter's article after the present Article was substantially completed. Like me, Professor Cooter suggests that efficiency analysis can be an ally of those attempting to fight discrimination, and that tax or other policies can be better and more efficient than regulatory approaches. There are, however, important differences between our approaches. Professor Cooter advocates transferable quotas or tax policies that require an ex ante decision as to participation rates by gender or race. My approach aims in part to avoid this aspect of social planning.

Having made this first part of my proposal, however, I want to immediately clarify it. First, I am *not* advocating, as Epstein does, a complete repeal of Title VII. Firms that categorically or irrationally discriminate against women would still be subject to penalties. The proposal is merely intended to permit firms an opportunity to justify, on exclusively rational grounds, lower pay for women. Indeed, firms will now have less reason not to hire women, *ex ante*, because they will no longer be able to point to the greater costs of women's labor. Second, I am only advocating this change in conjunction with the second prong, to be discussed below, of a significant change in the tax laws. The idea is to move the burden of the redistribution implicitly mandated by Title VII away from specific firms—the *good* firms, that in fact hire women—and onto the general population, mainly via higher taxes on married males. Absent this second change, the repeal of the equal pay laws may well, by itself, be *inefficient*, as I have discussed above.²⁰⁹ Finally, I readily concede that I have not taken into account the possible, and potentially very important, symbolic and psychological effects of the proposal. It may well be, for example, that receiving a lower salary, even for fully rational reasons and even if the tax system counteracts the pre-tax gap, is stigmatizing to women. If such factors are important, we may want to modify the proposal by continuing to require equal pay but giving a credit to firms that hire women, financed out of the general tax changes to be discussed below. All of these points are meant to underscore, once again, the tentative, illustrative nature of the proposal. Having stated a view of the problem, as one of involving not just objective indicia such as the gender gap but instead meaningful choice and respect for varying life plans, it becomes at once more important and more difficult to conceive of a solution.

In any event, the second part of the proposal is to change significantly the system of taxation, by following an optimal, or Ramsey, taxation scheme. This proposal, long supported in the public finance literature, would result in the government's taxing married men more, possibly much more, than married women. This change would alter the after-tax price schedule facing men and women, without placing the implicit redistributive burden on precisely those firms that were hiring women in the first place. It thus works hand-in-hand with the first prong to undo the market biases, however rationally grounded, against women.

I readily concede—indeed, I mean to emphasize—a certain tentativeness to the proposal. More studies are needed on precise behavioral effects to flesh out the details. There are also complex political, symbolic, tactical and even constitutional issues at stake.²¹⁰ But my primary interest is in generalities, and in showing how an efficiency-driven reform can be attractive on strictly

209. See *supra* Section IV.C.

210. I am currently exploring some of these issues in my research.

normative grounds. Section A gives some background on optimal tax theory, that I assume to be unfamiliar to most readers. Section B suggests how the tax-deregulatory proposal might interact with the model of Part III. Section C discusses normative justifications for the proposal, and Section D notes a few possible extensions.

A. *Optimal Tax Theory*

Because the reform proposal advanced in this Part depends critically on a move towards optimal tax treatment of men and women, this Section provides some general background on optimal tax theory. Problems of broad-based taxation typically throw us into a second-best setting, at least insofar as efficiency is concerned. Problems of preference-revelation make it virtually impossible to fund public goods with purely Paretian taxes that make everyone happy.²¹¹ It is in this setting that the perhaps mislabelled theory of optimal taxation, pioneered by Frank Ramsey (hence it is sometimes called Ramsey taxation) in the 1920's, arose.²¹² The idea of optimal taxation is that taxes should be designed to interfere as little as possible with aggregated individual choices, to minimize what is variously called deadweight loss or excess burden. An excess burden results from utility losses that are imposed by the distortion of a tax but not offset by revenue gain to the government. In other words, there is a pure loss in consumer surplus.²¹³

A simple example may help. Imagine that a consumer is willing to spend up to \$1.25 for an apple. Apples are currently priced, under competitive conditions, at \$1.00 each, so the apple buyer enjoys a surplus because she is getting something worth \$1.25 to her for only \$1.00. The government institutes a 50% tax on apples. In theory, the price of apples must eventually rise to \$1.50 because, under the competitive market assumption, firms were just breaking even when they charged \$1.00. In such a competitive market, the entire tax would have to be passed on to buyers in the long run.²¹⁴ The effect of this process in our example is that the apple buyer simply stops buying apples because their price now exceeds what she is willing to pay. There has been a deadweight loss here, because no one gains—not the consumer, who loses her excess utility; not apple farmers; not even the government, which

211. See MUELLER, *supra* note 149.

212. F.P. Ramsey, *A Contribution to the Theory of Taxation*, 37 ECON. J. 47 (1927). For elaborations on optimal tax theory, see David F. Bradford & Harvey S. Rosen, *The Optimal Taxation of Commodities and Income*, 66 AM. ECON. REV. (PAPERS & PROC.) 94 (1976); Agnar Sandmo, *A Note on the Structure of Optimal Taxation*, 64 AM. ECON. REV. 701 (1974).

213. MUSGRAVE & MUSGRAVE, *supra* note 195, at 277-95; Bradford & Rosen, *supra* note 212.

214. MUSGRAVE & MUSGRAVE, *supra* note 195, at 250-57. The example assumes a tax-exclusive tax, so that the 50% rate is imposed on the net price of the good, here \$1.00.

gets no tax.²¹⁵ The precise deadweight loss is measured by the sum of lost consumer and supplier surplus, offset by any revenue gain to the government.

The policy conclusions of Ramsey taxation are elegantly simple: *ceteris paribus*, taxes should be levied in inverse proportion to a good's (or activity's) elasticity.²¹⁶ Highly inelastic goods, such as medicinal drugs, cigarettes and perhaps gasoline, should bear a high rate of tax. Highly elastic goods, for

215. I assume no externality in the apples. Of course, a full analysis of the example would have to go into what happens with the \$1.00, and whether there are effects on the labor supply decision, etc. These are concerns that would move us into general equilibrium, whereas the example is pitched as one of partial equilibrium.

216. One qualification is that optimal tax policy tries to keep overall rates low. We can see this argument graphically. Figure 5 represents the familiar supply and demand curve, with inelastic supply, *s*, and a typical downward sloping demand curve, *d*.

FIGURE 5. *Supply and Demand with a Sales Tax*

When a simple per unit tax is put into effect, it shifts the supply curve up to *s'*, parallel to but above the prior supply curve by the amount of the tax. (Recall the textual example where the price of the apple rose by the amount of the tax; we are once again in a simple partial equilibrium setting.) In such a case, the deadweight loss or excess burden is given by the area of the triangle *ABC*; this represents the consumer surplus obtained at the prior equilibrium, which has now been lost. The general formula for the area of a triangle is given by:

$$A = (1/2) b h, \quad (1)$$

where *b* is the base and *h* is the height of the triangle. In this case, we have:

$$b = \Delta Q; h = \Delta P, \quad (2)$$

where ΔP and ΔQ represent the changes in price and quantity brought about by the tax change, respectively. But we know by hypothesis that

$$\Delta P = t, \quad (3)$$

The definition of price elasticity, *E*, yields:

$$\begin{aligned} E &= (\Delta Q/Q) / (\Delta P/P) \\ E &= (\Delta Q / \Delta P) P / Q \end{aligned} \quad (4)$$

Solving for ΔQ yields:

$$\Delta Q = \Delta P (Q/P) E. \quad (5)$$

But we know that ΔP equals *t*, so that Equation (1) now becomes

$$A = (1/2) (Q/P) t^2 E. \quad (6)$$

This tells us that the deadweight burden is proportional to the elasticity and the tax rate *squared*; it is thus not surprising that most of the optimal tax literature recommends low rates. See MUSGRAVE & MUSGRAVE, *supra* note 195, at 292-93; Joel Slemrod, *Optimal Taxation and Optimal Tax Systems*, 4 J. ECON. PERSP. 157 (1990); Stiglitz, *supra* note 160.

which ready substitutes exist—blue cars, for example, or Washington state apples—should bear low rates of tax.²¹⁷ The reason is clear enough: the less a tax changes behavior, the less inefficiency or deadweight loss. In the above example, imagine that our hypothetical consumer is completely inelastic when it comes to her daily apple. She will pay literally any price for her apple. She will still buy the apple even after the tax goes into effect and the price has risen to \$1.50. She has now lost fifty cents worth of utility—her surplus in any case is infinite, if she would literally pay any price for the apple—but the government has captured the fifty cents to pass on to someone else.²¹⁸

Two technical points follow. First, it should be clear that the philosophy underlying optimal taxation is utilitarian.²¹⁹ Taxes almost certainly make someone worse off, and the policy recommendation is designed to minimize the aggregate pain. Optimal taxes may even be Kaldor-Hicks inefficient, because the government can only hope to get exactly the money equivalent of a consumer's lost utility, as in the pure inelastic case. But if the wealth is transferred to a party or use with a higher marginal utility, the move will be utility enhancing. More fundamentally, optimal taxation is concerned with the choice of taxes. Given a fixed revenue need, optimal taxation seeks to find the set of taxes that will cause the least distortion. If we used uncompensated rather than compensated elasticities, this would lead directly to a Kaldor-Hicks efficient answer, as I show below.

This leads to the more complex second point. Traditional optimal taxation theory turns on what are called compensated, or Hicksian, elasticities.²²⁰ An uncompensated, or observed, elasticity has two components: income and substitution effects. The income effect arises because the tax lowers the individual's wealth, and this might change her preferences. The substitution effect arises because the taxed good or activity becomes less attractive relative to substitutes. In the case of a tax on wages, these effects are generally thought to work in opposite directions: (i) the tax makes one need to work more, in order to compensate for the loss of income, but (ii) the tax also makes leisure more attractive relative to work. Compensated elasticities are observed

217. Actually, the example indicates the importance of aggregating classes of goods. Apples may have a low elasticity, but Washington State apples a high one. In general, Ramsey taxation calls for general as compared to selective taxes. See Stemrod, *supra* note 216.

218. Once again, one needs to consider general equilibrium effects to be completely accurate. See generally McKenzie, *supra* note 161; Willig, *supra* note 161.

219. It is interesting to note that even Posner, who attempts to convert much of Kaldor-Hicks efficiency into ex ante compensated Pareto superiority, concedes that the move does not work relative to policies with substantial and nonrandom redistributive effects, such as the "substitut[ion of] a proportionate income tax for the current progressive one." See Posner, *supra* note 143, at 499.

220. See ANTHONY B. ATKINSON & JOSEPH E. STIGLITZ, LECTURES ON PUBLIC ECONOMICS 373 (1980) ("The importance in this formula of the *compensated* derivatives accords with intuition: the income effect would arise from any form of taxation, and the distortion stems from the substitution effect."); MUSGRAVE & MUSGRAVE, *supra* note 195, at 289-94.

elasticities corrected to take out the income effect. Thus, they only describe substitution effects, the exclusive concern of the optimal tax literature.²²¹

When we combine these two points—that optimal taxation is concerned with utility and that it looks only to compensated elasticities—an interesting further point emerges. Returning to the story of gender discrimination, we might be concerned with total, or uncompensated elasticities, on efficiency grounds. Just because utilitarians would only be concerned with substitution effects should not prevent us, with a different set of concerns, from looking at income effects as well. Why might we look to total elasticities, which combine income and substitution effects? There are at least two reasons.²²²

First, the Kaldor-Hicks concern for wealth would lead us to look only at uncompensated elasticities. This is because a concern for wealth is a concern for money, and not a concern for psychic utility. Uncompensated elasticities are to Kaldor-Hicks wealth maximization what compensated ones are to utilitarianism. This can be demonstrated by a simple example. Imagine a society with two potential taxpayers, Ms. Elastic (*E*) and Mr. Inelastic (*I*). *E* has an uncompensated, observed elasticity of one—this means that, for each one percent fall (rise) in her after-tax wages, *E* decreases (increases) her work effort by one percent. *I* has an uncompensated elasticity of zero—his work effort is completely unaffected by a change in his wage.²²³ This world has three states. In State I there are no taxes, and *E* and *I* each earn 100 units. In State II, the government imposes a tax of 50% on *E* only, cutting her after-tax wage, *ceteris paribus*, in half. Given her elasticity, *E* responds by cutting her work effort in half. The result is that *E* earns 50 units, 25 of which go to the government. The total social wealth is 150. In State III, the government places a 50% tax exclusively on *I*. Given his elasticity of zero, he responds with the same work effort as before the tax. Hence, *I* continues to earn 100, 50 of

221. An example may help clarify matters. Imagine taxing a good, such as gasoline, and then rebating exactly the amount of taxes we collected via some type of lump-sum transfers. Such a system would have only substitution effects. Each consumer would have the same amount of money she had before the tax was imposed. However, the relative after-tax price of gasoline would rise compared to all other goods. (Actually, there are some complexities here involving how the lump sum is calculated, but I put these aside). After the prices change, consumers would substitute other goods for gasoline, buying more fuel-efficient cars, ride-sharing, using mass transit, etc. Indeed, this is precisely the type of Paretian move wanted where externalities are involved, so that the market price is inaccurate, as may in fact be true for gasoline. Optimal taxation looks only to compensated elasticities because only the substitution effect is relevant to deadweight losses. The income effect, as exemplified in the inelastic apple or the gas tax, is a pure transfer—the type of redistribution countenanced by the second welfare theorem. The substitution effect, in contrast, alters individual behavior and thus distorts relative prices. See ATKINSON & STIGLITZ, *supra* note 220, at 373.

222. Actually, a third reason is the possible convergence between the compensated and uncompensated measures; this is the central idea of Willig, *supra* note 161. But cf. McKenzie, *supra* note 161.

223. It is interesting to note that some estimates of married male and married female elasticities in fact have men close to zero and women close to one, at least in the short term. See Hausman, *supra* note 74, at 238–49. For less directly related aspects of these issues, see Thomas MaCurdy, *Work Disincentive Effects of Taxes: A Reexamination of Some Evidence*, 82 AM. ECON. REV. 243, 248 (1992); George Zodrow, *Book Review*, 30 J. ECON. LITERATURE 916 (1992) (reviewing DO TAXES MATTER?, *supra* note 85).

which goes to the government. The total social wealth is 200. These numbers are summarized in Table 2.

State	<i>E</i> 's Net Wealth	<i>I</i> 's Net Wealth	Government Share	Total Social Wealth
I	100	100	0	200
II	25	100	25	150
III	100	50	50	200

TABLE 2. *Kaldor-Hicks Taxation*

It should be clear that State III is Kaldor-Hicks superior to State II. In State III, total wealth is greater, a necessary and sufficient condition for Kaldor-Hicks efficiency. It is easy to see that *E* could compensate *I* for his losses out of her winnings in moving from State II to III, and still be better off than she was in State II. States I and III are Kaldor-Hicks neutral, or noncomparable. The government is the winner in the move from I to III, and, were it to compensate the loser *I*, we would be back in State I. This illustrates that taxation can only be Kaldor-Hicks efficient if there are *negative* uncompensated elasticities. These occur when taxation leads to greater work effort, because income effects dominate substitution ones.²²⁴ Optimal tax, now redefined in Kaldor-Hicks terms, is once again concerned with the choice among taxing regimes. Here State III is preferred to State II. Note, finally, that this example has not been couched in terms of utility. Not only did I avoid using compensated elasticities, but I am also not factoring in the utility of money income. Depending on the relative utility functions, any of the States might be utility maximizing as compared to any of the others.

This example also helps to explain why statistical discrimination might be Kaldor-Hicks inefficient.²²⁵ If (holding the size of the groups equal) the group receiving lower wages by virtue of discrimination has a higher wage elasticity than that of the group getting the higher wages, there will be a net social loss, as the class getting the lower pay reacts more strongly by cutting back on its hours. (Indeed, if the higher-paid class has a *negative* elasticity, raising its wages may lead to a decrease in its hours worked, too.) This is exactly what occurs when the market separates married men and married women, as in Figure 3, above. Since married women have a higher wage

224. See, e.g., WALTER NICHOLSON, *MICROECONOMIC THEORY: BASIC PRINCIPLES AND EXTENSIONS* 656 (4th ed., 1989) (discussing backward-bending labor supply curves).

225. See Schwab, *supra* note 40, at 232-33.

elasticity, they respond to the lower wages firms offer them by disproportionately cutting back on their work effort, whereas men would be more likely to work despite lower wages. Similarly, when the market among women separates, as it does in the signaling equilibrium discussed in Part III, there is once again a Kaldor-Hicks inefficiency. The women least responsive to wage changes become educated and receive higher wages, leaving the remaining women to drop another notch in the pay scale and hence also in their participation. A similar dynamic seems to have developed among racial minorities under Title VII; the least elastic are making gains, while the more elastic are dropping out altogether. As much as one-fourth of the wage gains made by African-Americans relative to whites under Title VII seems to be due to the greater unemployment of African-American males.²²⁶ The same thing may be happening among women—gender gap gains are being fueled by the exit of the least committed—although the process is obscured by the supposedly “natural” pattern of women staying at home with their children.

A second reason to examine uncompensated elasticities returns more squarely to the model in Part III. Firms are seeing that men and women have different wage elasticities, and this is actually a part of the dynamic of rational discrimination. Taxing on the basis of total elasticities might move us from Ramsey to Pigouvian tax principles: the total elasticities may be playing a role in market failure. In the long run, the proposed tax change should affect the labor supply elasticities of men and women. By making men more elastic, and women less so, the change would impact the initial statistical discrimination against women. Firms would have a harder time differentiating men from women—not because they would have less information, but because behaviors would have converged. But this convergence will not be solely in the direction toward historically male patterns.

In any event, the policy ramifications of optimal taxation—generally in its traditional, compensated elasticity version—on the story of gender discrimination are rather straightforward, and have long been noted.²²⁷ Their advocates include Michael Boskin, former head of President Bush’s Council of Economic Advisors. The simple idea is that we should be taxing married women at lower rates, and married men at higher ones. To make such a proposal in nominally gender-neutral terms, and thereby to finesse constitutional norms,²²⁸ we should have a different tax schedule for secondary earn-

226. EPSTEIN, *supra* note 3, at 258 (citing James Heckman, *The Impact of Government on the Economic Status of Black Americans*, in *THE QUESTION OF DISCRIMINATION: RACIAL INEQUALITY IN THE U.S. LABOR MARKET* 50 (1989)).

227. See, e.g., Michael J. Boskin & Eytan Sheshinski, *Optimal Tax Treatment of the Family: Married Couples*, 20 J. PUB. ECON. 281 (1983); Daniel R. Feenberg & Harvey S. Rosen, *Alternative Tax Treatments of the Family: Simulation Methodology and Results*, in *BEHAVIORAL SIMULATION METHODS IN TAX POLICY ANALYSIS* 7-41 (Martin Feldstein ed., 1983).

228. Constitutional issues raised by “reverse discrimination” are beyond the scope of this Article. In general, facially neutral laws motivated by a desire to assist underprivileged groups may or may not be

ers, partially financed by higher taxes on primary earners. The financing would only have to be partial, because a lower tax on secondary earners would induce greater productivity and hence partially pay for itself.²²⁹ Such an approach would be proper under optimal tax norms for several reasons. By most accounts, primary earners have very low elasticities; they look like the committed apple buyer in the example above. Secondary earners, in contrast, are more elastic;²³⁰ they look like the first apple buyer, who gave up on apples in the face of the tax. These have been the primary actors in our story thus far, and their relative elasticities played a role in the market dynamics sketched out in Part III. But there is a third group, married women committed to staying at home. This group, too, is inelastic: no matter what the wage they face, their behavior remains unchanged. Optimal tax theory dictates that this group should be taxed highly because its behavior will be less subject to distortionary influence of the tax. In effect, raising the rates on primary earners while lowering them on secondary ones is an indirect means of addressing this third group.²³¹

In sum, optimal tax treatment would lead to taxing workers with elastic labor supply less than workers with inelastic supply. This means at a minimum that rates on married men would go up, and those on married women down. There are many practical means for approaching this goal that I do not discuss in detail here: returning to a system of individual income tax filing, perhaps with a more favorable schedule for the lower earner; re-instituting a second-earner deduction, or credit; changing the social security contribution system; allowing more favorable provisions for child care and other work-related expenses of the secondary earner.²³² Under virtually any efficiency or utilitarian standard, regardless of how they are implemented, tax policies ought

suspect, and it is an interesting question whether predicating the proposal on the relatively objective measures of optimal taxation is enough to overcome constitutional suspicions. *See generally* Mary E. Becker, *Obscuring the Struggle: Sex Discrimination, Social Security, and Stone, Seidman, Sunstein & Tushnet's Constitutional Law*, 89 COLUM. L. REV. 264 (1989) (discussing, inter alia, Supreme Court cases on gender discrimination under Social Security system).

229. This is like the case in the example of *E* and *I* above. For example, if the government in that case wanted to raise \$50 by an equal tax on both workers, the appropriate tax rate would be 29.28%. (This is the solution to a quadratic equation, formed by setting $50 = x(100 - x) + x(100)$, where x is the tax rate, and hence also the fall in *E*'s work effort.) To move towards the Ramsey result of a 50% tax on *I* only, the government would have to raise *I*'s tax rate less than it lowered *E*'s (20.72% compared to 29.28%).

230. Hausman, *supra* note 74; Lundberg & Startz, *supra* note 61.

231. It appears at first to be difficult in practice to tax so-called imputed income, and thus most proposals to tax domestic labor are quickly dismissed. However, the optimal tax proposed for the first two groups actually provides a means of optimally taxing this third group as well. Imagine giving a credit—a negative tax—for at least the first dollars contributed by a secondary spouse. We would then implicitly be taxing stay-at-home spouses, who would be forfeiting the benefit. The fact that at least some of the credit would be paid for out of higher taxes on primary earners further illustrates the idea. The aggregate (income) effect for households with two earners might be a wash: higher taxes on the husband would offset lower taxes on the wife. But where there is only one market worker, the higher taxes may be functionally assigned to the nonmarket worker. Since she, too, is inelastic, by hypothesis, this would be optimal in a Ramsey sense. For a similar point, see Posner, *supra* note 77.

232. I consider many of these ideas in McCaffery, *supra* note 78.

to change to favor (or, to put the matter better, to eliminate the bias against) married women qua secondary earners.

B. *Returning to the Model*

I now return to the basic tax-deregulatory reform proposal. Recall that this proposal combines greater contractual freedom in setting wages with a significantly altered set of tax laws. Relaxing the contractual bans will allow firms to cut whatever efficiency losses they now face due to Title VII; it would free firms to offer less money to women if this is the efficient thing to do, in a privately Kaldor-Hicks sense. Firms would have greater freedom in designing contracts to deal with the persistence problem.²³³ For example, firms might move more directly to gender specific wage-tenure profiles. Simultaneously, a higher tax on married men, and a lower tax on married women, not only comports with the general dictates of optimal taxation but also holds out some promise of unravelling the dynamics discussed in the model of Part III. If the deregulatory aspect of the reform leads to a greater pre-tax wage gap, the tax aspect will shrink the post-tax gap.

It is generally not possible to predict specifically, *ex ante*, all of the precise effects of this type of tax-deregulatory change. One of the ideas animating this entire inquiry has been that the present system has led to a restricted opportunity set for some workers. The idea is to open up the labor market so that more workers have more choices about more things; I want to avoid limiting married women to a stark choice between acting like men or staying at home. This openness necessarily means that I cannot predict, let alone dictate, what will happen when individuals are given fuller choices; that would be the type of paternalism I am attempting in part to avoid in relying on an efficiency norm. Genuine and respectful equality is not a matter of fitting all individuals into some predefined mold.

Despite this important qualification, I can predict that the tax and deregulatory changes would cause two general classes of effects. First, the greater contractual freedom from dropping the regulatory requirement of equal pay will increase productivity, if indeed the antidiscrimination laws have been mandating inefficiency, and will help women by making them more attractive to firms. If the firms do not have to pay equal wages to women, then women ought to become more attractive to firms.²³⁴ Women's pre-tax wages would

233. See Hersch & Reagan, *supra* note 57; Section III.B.2 and Figure 4 *supra* note 69 (discussing gender-specific wage tenure profiles). The tax change would flatten women's otherwise steeper profile by raising after-tax incomes at the lower end and by imposing higher rates on high earners, lowering the after-tax incomes at the upper end. This levelling off of the wage-tenure line might mitigate the normatively adverse self-selection effects by encouraging even those women uncertain of their likely tenure to enter the market.

234. I want to emphasize once again, as I did above, that I am only making this proposal in conjunction with a significant overhaul of the tax system; if we were to drop the regulatory regime alone,

drop, at least initially,²³⁵ and dynamic effects of the sort described in Part III would be set in motion. It is thus only by adding the tax changes to the proposal that the cycle can be broken; the regulatory change alone might be inefficient. Second, along with greater contractual freedom, the revised tax laws will mean that married women will face an after-tax wage schedule closer to their pre-tax one.²³⁶ For many women, the net effect of these two changes will be higher after-tax pay, which is arguably all that should matter to rational agents, since the taxes are coercively (and promptly, through payroll withholding) collected.²³⁷ At the same time, married men will see a fall in their real, after-tax wages.²³⁸ There will be a redistribution of resources from men to women, tied to the expansion of marketplace opportunities. But unlike the current Title VII mechanism, this redistribution will be spread across the entire economy, and not concentrated precisely on those firms that do hire women and pay them equally with men.

The basic intuition behind the proposal is easy to state. As noted above, the equal pay requirements of Title VII may be placing the burden in exactly the wrong place, both in terms of equity and efficiency. In contrast, taxes afford a means of changing the prices men and women face without directly impacting the pricing decisions of firms. The tax change would have two types of benefits. One, it would lessen the total distortionary effects of taxes; it would be Kaldor-Hicks efficient and utility enhancing, assuming, as the existing data supports, that both compensated and uncompensated elasticities point in the same direction for reform. Two, the change would shift the burdens of taxation from falling particularly heavily on women. This would presumably lead women to work more and men to work less, given a positive uncompensated elasticity for each. This alone would lead to a recalibration of employers' probabilistic assessments. The higher wage structure would also change the incentives to pursue education; the precise effects depend on the response of educational pursuit to the offered wage schedule. The new wage schedule might also remove some of the supply-side impediments—many aggravated by the tax system—to forming more vibrant part- and flexible-time work arrangements.

None of these effects would necessarily eliminate the gender gap. Indeed, it is possible (if a bit unlikely) that the effects would tend to destroy the sepa-

we would be squarely within the problems noted in Section IV.C. I am also ignoring the possibly important symbolic effects, consideration of which might lead to a change in the form of the proposal, as discussed above.

235. Note that if women's wages did not drop, then Title VII is probably having no adverse effect on the efficiency of the wage structure.

236. In the case of negative tax rates, women would face an after-tax wage schedule in excess of their pre-tax wage schedule.

237. Recall the current 50% tax rate on *lower* income wives. The higher real after-tax wage might bring many back into the work force and impact the human capital decisions of others.

238. The fall should be lessened by productivity gains and, given their elasticities, the relatively smaller increase in their marginal tax rates, but one would expect a net fall nonetheless.

rating equilibrium sketched above—that is, between committed women acting like men and uncommitted women staying at home—and leave us with a pooling equilibrium that lumps all women together, to the detriment of the truly committed group.²³⁹ Additionally, problems of the second-best, combined with the complex sociopolitical questions regarding the effects and propriety of implicitly gender-based tax schedules, make my policy proposals extremely tentative.²⁴⁰ But one issue that I want to emphasize, at a theoretical level, is that under an efficiency-driven analysis closing the gender gap is not a primary or even a direct goal; respecting individual choices is. The existence of a gender gap where men and women are equally productive is *evidence* of inefficiency. Such a gap violates the productive efficiency conditions that would obtain under ideal market conditions, justifying the type of scrutiny employed in Part III. In any perfect market outcome, each worker would be paid her marginal productivity at all times. We can aim to move toward that state, but we cannot presume that women and men will make the same choices regarding education, labor market participation, and otherwise that affect their productivity in the first place. A central idea in relying on efficiency norms is that we should respect individual choices, whatever they are.

C. *The Normative Propriety of the Proposed Solution*

This subsection advances three distinct but related arguments for the propriety of the proposed tax-deregulatory solution in an admittedly second-best setting. These are not discussed in any special order of priority. First, a utility-enhancing and/or Kaldor-Hicks efficient move, both of which the proposal seems to be, is generally a good second-best solution even on liberal grounds, where perfect market outcomes are not forthcoming. As a second-best solution, however, such a move is subject to normative (e.g., rights-based) constraints. No such constraints are made out in this case, where the principles have led us to favor a historically disfavored group. In other words, utility enhancements are generally to be preferred where some government action is necessary and such action will inevitably generate some winners and some losers.²⁴¹ This will be true where, as in our model, there is some exogenous

239. Even this might be efficient in a Kaldor-Hicks sense; recall Stewart Schwab's argument that rational discrimination is inefficient if the lower-paid group has a more elastic labor supply, as we might have here. See Schwab, *supra* note 40, at 230; see also RASMUSEN, *supra* note 45, at 205-11 (demonstrating, inter alia, that pooling can be second-best efficient).

240. I readily concede, for example, that there are difficult unresolved questions about the pace of change and the nature of transitional relief, if any. See generally Stephen Coate & Glenn Loury, *Antidiscrimination Enforcement and the Problem of Patronization*, 83 AM. ECON. REV. 92 (1993) (suggesting gradual antidiscrimination programs are preferable to abrupt ones); Martin Feldstein, *On the Theory of Tax Reform*, 6 J. PUB. ECON. 77 (1976) (discussing, inter alia, differences between optimal tax reform and optimal de novo design, with attention to transitional problems involved).

241. Cf. Posner, *supra* note 143, at 499.

revenue need, but no readily available Paretian means of meeting it; some people will pay more than society is giving them back, and others will pay less.

Within a liberal argument, however, the case for the reform proposal cannot be grounded on utility or wealth maximization alone. Instead, it must rest on something like the fact of minimal interference with private preferences. Recall that this is precisely what optimal tax theory aims to do: minimize the total behavioral distortions of tax.²⁴² Optimal tax theory, which evolves out of a utility-maximizing social welfare model, is in fact also about respecting individual preferences. Optimal tax is fundamentally concerned with minimizing the aggregate effects of taxation on individual choice. Whether coincidentally or not, this is precisely one of the concerns in the case of gender discrimination—the greater burden imposed on women's choices by the current tax regime and other market failures. The argument for deregulation is similarly based on reducing distortions. There is a strong *prima facie* appeal to the claim that the tax-deregulatory proposal is the best solution in a second-best world: given that we have to veer away from the behavioral attributes of purely consensual mechanisms, such as Paretian efficiency conditions, we ought to do so as little as possible, in some meaningful sense of “as little as possible.”

Precisely because we have deviated from purely consensual, Paretian ethics, however, the *prima facie* case is not enough on liberal grounds. We are harming some individuals; within a liberal framework this raises important normative questions. The first argument thus needs another step: once we have determined that Paretian taxes are not forthcoming, and have developed an optimal tax solution, we next need to scrutinize that solution to see if there are compelling normative objections to it. For example, optimal tax may lead us to tax life-sustaining drugs, because the demand for these will be highly inelastic. However, such a tax would almost certainly violate social norms of fairness.²⁴³ In the case of wage taxation, optimal tax leads us to taxing secondary earners much less than primary ones. Will this run into normative constraints? My argument is that it will not. The burdens of the change will fall on married men and traditional single-earner families, and the benefits will

242. Cf. Sandmo, *supra* note 212 (arguing that optimal taxation is concerned with making the best of a necessarily distortionary tax system).

243. A fuller discussion of this example is helpful. Imagine that we were to separate the revenue-raising and expenditure functions, and to pursue optimal tax norms for the former. Then we would indeed be led to tax life-sustaining drugs. However, it is highly likely that this tax, in and of itself, would create an equitable claim for wealth transfers to those who needed the drugs. Our revenue needs would then increase, and we would have to engage in further taxation. Ultimately, we would have to tax individuals other than those needing the drugs, or the cycle would not close. Especially in a general equilibrium setting, taxing the drugs is likely to cause efficiency losses. So this is just another way of saying that the principles of optimal tax, within a pluralist-liberal society, must yield to normative concerns.

flow to married women (perhaps all women) and two-earner families. This is a shift from a historically privileged to a historically underprivileged group.

This discussion shows why the analysis of Part III is relevant to the liberal efficiency case, while it would not be needed under a Kaldor-Hicks or utilitarian approach, both of which directly support the proposal. Inefficiencies under liberalism become just like any other injustice: they mandate scrutiny, and if a particular group can make out a case for being harmed, and a solution that appears to make things better—in the sense of being at least close to what would emerge under perfect market conditions—is forthcoming, we should pursue it. Even without the complete analysis of Part III, it is easy to see that the failure to follow optimal tax norms all along has led to a severe and perverse cycle against women: it has added a tax disincentive to their already greater labor supply elasticity, compounded by, and compounding, statistical discrimination against them.

It is possible to restate this argument in neutrality terms, along the lines I pursued above. Markets and their efficiency properties appear to be neutral, in that efficiency can be justified without reference to its effects on particular groups or visions of the good life. Given market failures, however, this *prima facie* neutrality does not exist. As in the tax example discussed above, where the non-neutrality of head taxes led to a change, flawed market processes and outcomes are not neutral: they affect identifiable groups or visions of the good life in distinct ways.²⁴⁴ So long as we believe in the general neutrality of perfectly functioning markets, it is not necessary to make a frontal, Marxist-type or other assault on the morality of markets. Market failures become a grounds for opening up liberal dialogue about constructing a more neutral social order. If a market failure can be shown to affect a given group and some change, responsive to the failure, is apparent, efficient and utility-enhancing, and solicitous of the very group being harmed by the market failure, why should we not pursue the reform?

Further, the optimal tax solution can be seen in fact as more neutral than the status quo. Whereas existing tax laws insist on the static, nominally neutral norm of taxing "equal earning couples equally,"²⁴⁵ in reality the present law has distinctly non-neutral effects. Since secondary earners are more elastic, perhaps due to the very social conditions and market failures discussed in Part III, the intended neutrality of the tax law in fact imposes a greater real burden on them. Optimal tax theory aims for neutrality in the imposition of burdens; it takes into account the urgency of preferences.

The second argument is that the optimal tax solution, in this particular case, actually looks like a Pigouvian or market-correcting tax. The problem of

244. It is also possible to pose this neutrality argument, once again, in equality terms, by noting that the existing market structure is not affording women and men, or one- and two-earner couples, an equal opportunity to pursue their life plans.

245. See McCaffery, *supra* note 78, at 986-87, 1035-58.

greater secondary earner elasticity was centrally connected to the informational cycles discussed in Part III; it was both a cause and an effect in the feedback loop leading to discrimination against married women. It was a cause because the assumption of lesser female persistence or commitment set in motion the downward spiral against all women. It was an effect because women found themselves being separated into groups based at least in part on their elasticities, and because the lower after-tax salaries they confronted affected elasticities.²⁴⁶ In other words, the social cost of the market imperfections is directly related to elasticities, so that taxing on the basis of elasticities looks like taxing other sources of market failure, as in the classic case of externalities.²⁴⁷

This argument is conceptually distinct from the first one, that advocated making a utilitarian or wealth-maximizing decision, subject to unspecified normative constraints. In this case, I am specifically looking at a source of market failure, and aiming to correct it. Once again, there is the problem of second-best—I cannot say that the tax-deregulatory solution will restore prices to some free market level, since it will not remove all market failures. Indeed, any remaining taxes will continue to create market distortions. It is important to note, however, that the proposal turns on changing the nature of an existing market failure, rather than intervening in a new domain of transactions. Thus the categorical antimarket intervention arguments are not appropriate here; there is no reason to fear the general parade of horrors that Epstein and others drag out as an argument against all regulation.²⁴⁸

Still, the proposed change would have to be justified by some combination of empirical assessment, normative judgment, aesthetic predilection, and pure hunch.²⁴⁹ In rejecting utilitarianism, we cannot accept that a theoretical welfare gain will be decisive. We need to consider the failures sketched out in Part III and ask whether a change in the tax schedules, combined with relaxing the regulatory injunction for equal pay, will make the labor force more open to accommodating individual preferences than it is now, and whether this greater accommodation moves that market in a normatively attractive direction. These thoughts begin to merge into the third argument for the change.

246. That is, we could presume that elasticities are endogenous. Workers will have lower elasticities, and nonworkers higher ones. Even assuming that preferences do not change (see George Stigler & Gary Becker, *De Gustibus Non Est Disputandum*, 67 AM. ECON. REV. 76 (1977)) elasticities would still vary along income levels. A worker facing a low paying job might have a more elastic response to changes in her wage than a highly paid worker would to the same price change. See NICHOLSON, *supra* note 224, at 195.

247. See, e.g., Baumol, *supra* note 83.

248. To be sure, difficult decisions will have to be made regarding setting the rate schedules, but there need be no standing bureaucracy challenging firm decisions, overseen by a judicial system and so on.

249. Rawls, for example, frequently makes reference to hunches in forming the details of a political program. For example, in speaking of limiting unequal distributions of wealth, he states: "The taxes and enactments of the distribution branch are to prevent this limit from being exceeded. Naturally, where this limit lies is a matter of political judgment guided by theory, good sense, and plain hunch, at least within a wide range." RAWLS, *supra* note 160, at 278.

This third argument is that the only way to go about efficiently changing the dynamics of Part III is to alter the incentive structure confronting women. The status quo is not working. Mere distributive policies, as I noted in Section IV.D, will not be sufficient to bring us closer to a normatively attractive efficiency. I do not necessarily need to make the case for some change on the prior grounds of rights or justice, as by arguing that women have a right to equal opportunity that existing market structures impede, although such arguments would indeed be relevant, and possibly decisive, here. Instead, I can simply look to the analysis of Part III to conclude that there is significant inefficiency at work. The third argument is that only a change that alters the opportunity sets of women in the job market can even hope to unravel the dynamic vis-à-vis women. This appears to be part of the theory behind regulatory interventions such as Title VII, at least insofar as they have efficiency justifications.²⁵⁰ What the third argument aims to do is to alter the predicates on which firms' rationality is based.

This third argument reverses the claim from efficiency, and changes the structure of the argument. It assumes in the first instance a desire to benefit women—although, as I noted above, this desire can itself be rooted in liberal efficiency norms—and then asks how this can be done efficiently.²⁵¹ In general, this argument comports with the structure of the second welfare theorem, in that it is an attempt to change initial endowments, and then let market forces take over. The problem with the second welfare theorem, at least as commentators such as Epstein explicitly use it, is that it defines “endowments” too narrowly. Endowments are relevant in market theory because they shape the contours of opportunity sets. In a hypothetical marketplace, for example, some will bring apples, others will bring oranges, widgets, or money, and so on, into the market. There is nothing sacred or special about any such goods. What is important is that the value the market places on the goods delimits the market opportunities available to each trader. People maximize over their opportunity sets, which happen to be defined with reference to their initial endowments. It is these opportunity sets, and not the typically physical, material resources that define them, that ultimately matter. Very often, giving people money or other resources is the best way to expand their opportunities. But the distinction between opportunities themselves and the goods that generate them is relevant where, as in the case of gender discrimination, there are problems arising from market failures in translating endowments into opportunities. Once again, a different perspective emerges by looking at the issue from the perspective of equality norms.

250. See, e.g., Donohue, *Economic Perspective*, *supra* note 4.

251. See Lawson, *supra* note 143; cf. discussion regarding efficiency, *supra* Sections III.A-B; Cooter, *supra* note 208 (setting out a similar task for himself).

Women do not necessarily need more money. They do not necessarily need more education. What they need is a market more accepting of their choices, without their having to contort their behavior to fit historically male labor-market participation patterns. Epstein's basic point is that we cannot get these expanded opportunities by fiat, because we cannot expect firms to act inefficiently.²⁵² Indeed, we can make this argument even stronger. The terms of traditional regulatory intervention are themselves influenced by a patriarchic social order, so that the antidiscrimination laws may even be consciously trying to squeeze women into a male pattern of work and family life—Title VII may actually be a cause of the paradox of better paid but less happy women. Like the well-meaning but ultimately wrong-headed protectionist legislation dating from the first part of this century,²⁵³ Title VII may have it all wrong. First, it defines the terms of equality, then it pushes people toward it. This is a strange way to go about respecting individuals.

In contrast, we can change the incentive structure for women by changing our tax laws. That at the same time we would be throwing some barriers in the way of men's ultra-commitment to the market will serve a similar end. These changes would not ask firms to act irrationally; they would instead change the incentive structure facing employees. This move appears to be more promising than either traditional regulatory or redistributive policies. This third argument looks like a subjective efficiency claim. The only way to help women, efficiently, is through a style of intervention properly solicitous of both efficiency and market failure concerns.

In concluding this subsection, I want to stress that the three styles of argument are quite distinct. The first one (provisionally accepting efficient, utility-enhancing social change) leads to a utilitarian or wealth-maximizing approach, subject to unspecified normative constraints. The second (Pigouvian or market-correcting taxation) looks for evidence of market failure directly and seeks to correct for it. The third (efficient means of serving antidiscrimination goals) seeks the best way of expanding opportunity sets in any given context. Quite clearly, these three arguments may point in different directions, in the case of gender discrimination or otherwise. For example, the first might look to compensated elasticities, whereas the second would look only to uncompensated ones; the third argument would not care about elasticities at all, except as they might be evidence of a market failure, justifying the intervention on efficiency-based grounds. That the tax-deregulatory proposal is justified by the rough convergence of all three arguments ought therefore to weigh especially in its favor.

252. This is the specter of rent control again; we cannot set categorical rules and expect that all will work out, when we know full well that firms (or landlords and other tenants) will continue to act rationally in their self-interest.

253. See GOLDIN, *supra* note 2, at 190-91.

D. *Extensions*

This Article has undertaken an extended look at gender discrimination in the work force, and an exploration of the relevance of efficiency norms for overcoming cycles of entrenched bias. Because the project involved numerous contextual decisions and compromises wrought from the second-best setting, it is difficult to extend its conclusions as is to other areas of law and society. I do believe, however, that tax-oriented solutions offer great promise in other areas, and in a wide range of cases may be preferred to either regulatory or purely redistributive solutions. In each case, however, the recommendation must be tempered with an understanding of the particular effects of market failure, and the normative analysis of the prior subsection may lead it in different directions.

Consider, for example, the case of racial discrimination. Epstein would eliminate employment discrimination laws here, too, allowing rational discrimination to continue.²⁵⁴ But much of the analysis of gender discrimination is relevant here as well. We are not in a first-best setting, because transaction costs have led to a salary structure that cannot correspond to marginal productivity at all points. Firms must make guesses as to ability, persistence, and other important attributes. The guesses will be based on existing data, but will then form a basis for the incentive structure. Racial minorities will be paid less, because at some prior point they may have been thought to pose greater risks or be less productive, perhaps based on statistical testing that has its own racial biases, or because of a denial of access to higher education, and so on.²⁵⁵ A cycle similar to that observed with respect to gender will ensue.²⁵⁶ The market will separate, with some potential workers pursuing education and other proof of quality and commitment, and others dropping out altogether. As Epstein has argued, traditional regulatory remedies may even make these cycles worse. For example, the mere knowledge of the existence of affirmative action programs may undercut education's signaling value for many minorities.²⁵⁷ Unlike white women, minorities may have no way out of the vicious cycle because education cannot help them as much as it helps women. Racial minorities may be left either accepting lower-paying jobs or withdrawing from the work force altogether. Firms will make probabilistic updates and the cycle will continue. As in the case of women, progress among the most committed minorities may mask the underlying problems.

254. EPSTEIN, *supra* note 3, at 242-66.

255. *See, e.g.,* Kelman, *supra* note 31.

256. *See, e.g.,* Bulow & Summers, *supra* note 46.

257. *See* STEPHEN CARTER, REFLECTIONS OF AN AFFIRMATIVE ACTION BABY (1991); Coate & Loury, *supra* note 61, at 13-21.

In the face of such cycles, brought about at least in part and exacerbated in any event by market failures, there is no compelling reason to embrace the do-nothing approach on efficiency grounds. Merely insisting on Pareto superior moves, commencing at an arbitrarily chosen date, will lack much normative appeal. Redistribution alone will not help: it will not change any of the incentives or opportunities facing minorities, and may even make these worse. A vicious cycle of poverty and dependence will ensue. Once again, what is needed is an enhanced opportunity set.

A difference between minorities and women is that the optimal tax solution per se may not be forthcoming. Indeed, some data indicate that those minorities who are working are rather inelastic; presumably, the income effect is strong.²⁵⁸ Thus, the second two arguments developed in the prior subsection may become paramount. First, a more favorable tax rate on members of racial minorities would be needed as a corrective to market failure; second, it is the best way to alleviate cycles of discrimination and poverty for all minority individuals, while allowing firms to act rationally in a market setting. Indeed, the tax idea may comport with the neutrality norm of equal sacrifice; a member of an underprivileged minority has had to work harder for the same amount of wealth, *ceteris paribus*, and would sacrifice more in real terms by giving up a like nominal amount. These ideas of course need development, perhaps once again in conjunction with a change in the existing regulatory approach to antidiscrimination.

Such a proposal is not far-fetched. Indeed, something of its spirit lies behind the earned income credit,²⁵⁹ designed to give a negative tax rate to lower earners and thus to supplement the rational market-pricing signal. I believe that there are some problems with the earned income credit—its inordinate complexity, its creation of high marginal rates on lower middle income earners, its undue burden on marriage, to name three²⁶⁰—but it seems to be a move in the right direction. The proposal here would go even further, to allow members of underprivileged minorities to claim more favorable tax rates. Such a proposal would likely be more controversial than a benefit for secondary earners, and would no doubt involve more intricate constitutional scrutiny. I thus mean only to suggest it as an avenue for further consideration, and as another illustration of the general approach set forth here.

VI. CONCLUSION

I have tried to advance three sets of arguments in the three major Parts of this Article. In Part III, I used simple models to show how gender

258. See Hausman, *supra* note 74, at 253-60.

259. I.R.C. § 32 (1993).

260. See McCaffery, *supra* note 78, at 995-96, and sources cited therein.

discrimination in the labor market is not efficient in any normatively appropriate sense of that term. Instead, much of the gender status of the work force can be understood as the result of market failures, playing themselves out over time. These failures began with transaction costs that kept employees from getting simple contracts that paid them their marginal productivity at each point in time. From there, imperfect and asymmetric information, taxes, and incomplete markets led to various distortions. Over time, women in the labor force separated into two groups: those that withdrew from the work force, and those that became highly committed to it. Within this setting, the observed narrowing of the gender gap is neither surprising nor necessarily cause for celebration. Because of the market failures, women have been given a stark choice: act like men have traditionally acted in the work force, or get out. The market, deeply flawed, has been dictating choices rather than accommodating them, as liberal theory would have it. Women have been slouching towards equality on terms and conditions laid out in advance, and carried over from a patriarchic world. This helps to explain the coexistence of objective measures of improvement and subjective measures of despair.

In Part IV, I criticized the do-nothing approach advocated by Richard Epstein and others. I argued that this approach rests at best on a negative version of Pareto superiority, justifiable only under certain libertarian norms. In contrast, utilitarianism, Kaldor-Hicks wealth maximization, and a liberal sense of efficiency all point to the need for change. There is no very good argument that doing nothing—as by having no regulatory antidiscrimination regime—will move society any closer to what is normatively appealing in an efficiency norm. At the same time, I argued against mere redistribution, which would seek to right past wrongs by transferring money or other material resources to harmed parties. In the context of gender discrimination in the work force, where the incentive structure for women has been diminished, money alone will not solve the problems, and may well make them worse. I also discussed a theme carried throughout the Article—that the gender gap is not an especially good measure of equality, because it fails to get at what is truly important, such as our equal capacities or opportunities to pursue varying life plans and projects.

Finally, I argued in Part V that a liberal commitment to efficiency would lead to a policy of intervention, and that tax policies, as opposed to regulatory or narrowly redistributive ones, may be the preferred means to combat gender discrimination. My conclusions here were necessarily tentative, because of the problems of second-best, numerous empirical questions, and the complex sociopolitical issues that a complete programmatic solution inevitably poses. I nonetheless pursued a set of policy proposals, at least illustratively. I conceded Epstein and Posner's case against the equal pay aspects of the antidiscrimination regime, but proposed a major overhaul of the tax system to supplement the relaxation of the regulatory rules. I sketched out the optimal

tax approach, which would lead to higher taxes on primary workers, typically married men, and lower taxes on secondary ones, typically married women. I discussed how such an approach could play a role in unravelling the dynamic sketched in Part III. I argued that there were compelling reasons, under Kaldor-Hicks, utilitarian, and liberal efficiency norms, to accept some means of taxing primary more than secondary earners as a particular, partial solution to gender discrimination in the work force. I noted very briefly that similar approaches might work in other areas of discrimination, such as racial discrimination.

In addition to these three distinct sets of arguments, I developed a unifying set of themes throughout the Article. First, the efficiency of perfect markets generates a vision with much appeal on liberal grounds; it represents a certain type of neutrality, and a realization of the notion that society respects individual autonomy and preferences. But it is critical to sort through the competing, divergent senses of efficiency and their underlying norms in order to make coherent and compelling efficiency claims in the real world of imperfect markets.

Second, in contrast to theory, the story of the real world is one in which market failures dominate, and have profound effects over time. We must be careful in drawing conclusions from first-best theory, or on the basis of certain limited objective measures, such as the gender gap, alone. Market failure is dangerous precisely because it distorts individual choice and limits individual opportunities. These are distortions that, over time, can be severe. This analysis has important implications for equality talk, as well. Meaningful, respectful equality can only be obtained when we look at opportunity sets and respect the diverse choices that people might make. There is no reason to push everyone into the same mold.

Finally, market failures do not give us a reason to abandon efficiency as a normative ideal. While market failures put us into the difficult and contestable region of second-best theory, efficiency remains an attractive goal because it respects traditional liberal principles such as decentralization, fairness, neutrality, and individual autonomy. Efficiency-driven analysis is powerful because it can reveal problems and solutions. Just because efficiency-related normative prescription turns out to be as much art as science is hardly grounds for abandoning it altogether, or for letting the particular distortions of the present dictate, forever, the terms of the future.